SEMI-ALGEBRAIC TECHNIQUES IN VARIATIONAL ANALYSIS: PSEUDOSPECTRA, ROBUSTNESS, GENERIC CONTINUITY, AND MOUNTAIN PASS ALGORITHMS.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jeffrey Pang Chin How

August 2009

SEMI-ALGEBRAIC TECHNIQUES IN VARIATIONAL ANALYSIS: PSEUDOSPECTRA, ROBUSTNESS, GENERIC CONTINUITY, AND MOUNTAIN PASS ALGORITHMS.

Jeffrey Pang Chin How, Ph.D.

Cornell University 2009

Variational Analysis is the modern theory of nonsmooth, nonconvex analysis built on the theory of convex and smooth optimization. While the general theory needs to handle pathologies, functions and sets appearing in applications are typically structured. Semi-algebraic functions and sets eliminates much of the pathological behavior, and still forms a broad class of constructs appearing in practice, making it an ideal setting for practical variational analysis. Chapter 1 is an introduction to the thesis and Chapter 2 reviews preliminaries.

Chapters 3 to 5 describe various semi-algebraic techniques in variational analysis. Chapter 3 gives equivalent conditions for the Lipschitz continuity of pseudospectra in the set-valued sense. As corollaries, we give formulas for the Lipschitz constants of the pseudospectra, pseudospectral abscissa and pseudospectral radius. We also study critical points of the resolvent function. Chapter 4 studies robust solutions of an optimization problem using the "ϵ-robust regularization" of a function, and prove that it is Lipschitz at a point for all small $\epsilon > 0$ for nice functions, and in particular semi-algebraic functions. This result generalizes some of the ideas in Chapter 3. Chapter 5 studies the continuity properties of set-valued maps. We prove that the set of points where a closed-valued semi-algebraic (and more generally, tame) set-valued map is not strictly continuous is a set of lower dimension. As a by product of our analysis, we prove a Sard-type theorem for local (Pareto)

minimums of scalar-valued and vector-valued functions.

Chapter 6 departs from the theoretical bent of the rest of the thesis and is computational. We describe algorithms for computing critical points of mountain pass type. We prove that sub-level sets of a function coalesce at critical points of mountain pass type and discuss algorithmic implications. In particular, we propose a locally superlinearly convergent algorithm for smooth nondegenerate critical points of Morse index 1. We conclude this chapter by describing a strategy for the Wilkinson problem of finding the closest matrix with repeated eigenvalues.

BIOGRAPHICAL SKETCH

Chin How Jeffrey Pang was born in Singapore in 1979. He had his pre-university education in Singapore, and represented Singapore in the International Mathematical Olympiads in 1995 (Canada), 1996 (India) and 1997 (Argentina). He graduated with a Bachelor degree in Mathematics with Honours from the National University of Singapore (NUS) in July 2003, and a Masters in the Singapore MIT Alliance in July 2004.

He came to Cornell University in August 2004 to study for a PhD in Applied Mathematics under the direction of Adrian S. Lewis, and has been in Cornell for five years. In Fall 2008, he was a visiting scholar hosted by Aris Daniilidis at the Centre de Recerca Mathemàtica, Barcelona. In Fall 2009, he will be at the Fields Institute based in the University of Toronto as the Marsden Postdoctoral Fellow for the thematic program on the Foundations of Computational Mathematics. In Spring 2010, he will continue his postdoctoral work at the University of Waterloo.

To all who have taught me Math

ACKNOWLEDGEMENTS

First and foremost, I wish to thank my adviser Adrian Lewis for his guidance throughout my PhD program. I benefitted much from his wisdom. I gratefully acknowledge him for introducing me to the field of variational analysis, for coaching me on all aspects of being a better mathematician, and for his financial support throughout the PhD program.

Secondly, I thank Aris Daniilidis for hosting me at the Centre de Recerca Mathemàtica, Barcelona in Fall 2008. I had a wonderful time working with him and chatting with him about things mathematical or otherwise. The material in Chapter 5 is a result of our work during that time.

I would like to thank my committee members, James Renegar and Charles Van Loan, for cheerfully agreeing to serve on my thesis committee and for the excellent classes I took with them. I thank the other people working in continuous optimization in Cornell for the pleasant academic environment in Cornell. In particular, I thank Mike Todd for his insightful comments (one of which led to a sharpening of a result in Chapter 4), and I thank Stefan Wild and Dennis Leventhal for organizing the continuous optimization seminars. I thank the director of CAM, Steve Strogatz, and the Admin Manager of CAM, Dolores Pendell, for keeping CAM functioning on a daily basis.

I thank many of the professors in the National University of Singapore for guiding me through my Math education, especially those since the days I was involved in training for the IMO. I am especially grateful to those who gave me a chance to try out for the IMO team back then, even when it looked like I never had any chance at that time.

This thesis was typed using LaTeX and LyX. Thanks to their creaters.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Variational analysis is the modern theory of nonsmooth, nonconvex analysis built on the theory of convex and smooth optimization – see the text by Rockafellar-Wets [80] and others, for example, Aubin-Frankowska [7], Borwein-Zhu [16], Clarke [31], Clarke-Ledyaev-Stern-Wolenski [32], and Mordukhovich [72]. While the general theory of variational analysis needs to accommodate pathological examples, constructions typically appearing in practice enjoy some of the properties of smooth functions and convex functions. For this reason, a good part of variational analysis focuses on favorable classes of nonsmooth, nonconvex functions, like Clarke regular functions, prox-regular functions and amenable functions.

A recent direction in variational analysis is the study of variational properties of semi-algebraic functions and sets. Semi-algebraic objects eliminate most of the pathologies in analysis, and still form a broad class of objects appearing in practice. Semi-algebraic objects are defined by a finite set of polynomials, and include, for example, piecewise polynomials, rational functions, and the mapping to eigenvalues. The pseudospectral mapping, which is the focus of Chapters 3 and 6, is semi-algebraic. The set of positive semidefinite matrices, a common set in optimization, is semi-algebraic.

Semi-algebraicity in concrete problems can often be easily checked using the Tarski-Seidenberg principle. Semi-algebraic geometry can be studied in an axiomatic manner under the broader theory of definable functions and sets, and o-minimal structures. O-minimal structures enjoy a variety of favorable properties which are employed in parts of this thesis, like the local conic homeomorphism and various decomposition theorems. We can go further and formulate results

for tame sets, which are sets whose intersection with every ball is in some o-minimal structure. Techniques in semi-algebraic geometry, o-minimal structures and tame topology are now applied in various problems in variational analysis. See for instance [5] (convergence of proximal algorithm), [13] (Łojasiewicz gradient inequality), [14] (semismoothness), [52] (Sard-Smale type result for critical values) or [53] for a recent survey of what is nowadays called *tame optimization*. Section 2.2 provides an extended introduction to semi-algebraic geometry.

Much of my work in variational analysis is on set-valued maps. We say S is a *set-valued map* from X to Y, denoted by $S : X \rightrightarrows Y$, if for every $x \in X$, $S(x)$ is a subset of Y. Many problems in applied mathematics are inherently set-valued in nature (for example, problems in feasibility and control), and are often best treated with set-valued maps by appealing to set-valued analytic tools like continuity concepts and chain rules. In variational analysis, the generalized gradients of functions and the tangent and normal cones of sets are important set-valued maps.

Chapters 3 to 5 describe some of my work on semi-algebraic variational analysis. Chapter 3 illustrates an application of variational analysis in the study of pseudospectra. Chapter 4 illustrates how semi-algebraic variational analysis techniques can be applied to a general theory of finding robust solutions to optimization problems. Chapter 5 studies the continuity properties of semi-algebraic, and more generally tame, set-valued maps. The next few sections of this chapter give a more detailed introduction.

Chapter 6 departs from the theoretical nature of the preceeding chapters, focusing instead on numerical methods for finding critical points of mountain pass type. This work was motivated from the work in Chapter 3 which relates the Lip-

schitz continuity of the pseudospectrum to the avoidance of (nonsmooth) critical points. In the smooth case, critical points are where the derivatives vanish. While critical points that are maximizers or minimizers can be found using optimization, one uses a mountain pass algorithm or its variants to find critical points that are neither maximizers nor minimizers. Recent work by Alam and Bora [1] reduced the Wilkinson problem of finding the distance of a matrix A to the closest matrix with repeated eigenvalues to finding the lowest saddle point of mountain pass type that connects two components of the pseudospectrum, referring to such saddle points as "points of coalescence". This perspective gives promising numerical results, and does not seem to be well studied previously.

Chapter 2 recalls some of the preliminaries that will be used often throughout the rest of the thesis. The rest of the sections in this chapter gives an introduction to the respective chapters. The material in each chapter is based on [62, 63, 35, 64] respectively.

1.1 Variational Analysis of Pseudospectra

My research started from finding conditions for the Lipschitz continuity of pseudospectra. The reader will be familiar with eigenvalues and their applications in applied mathematics. As we consider perturbations to an $n \times n$ complex matrix A with spectrum $\Lambda(A)$, we are led to study the ϵ-pseudospectrum, $\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$, which is a set-valued map defined by:

$$\Lambda_\epsilon(A) \;=\; \{z \mid \exists E \in M^n \text{ such that } \|E\| \leq \epsilon, z \in \Lambda(A+E)\},$$

where M^n is the space of matrices of size $n \times n$. A well-known equivalent formulation, assuming $\|\cdot\| = \|\cdot\|_2$ as we do throughout, is

$$\Lambda_\epsilon(A) \;=\; \{z \mid \underline{\sigma}(A - zI) \leq \epsilon\}$$

where $\underline{\sigma}(A)$ denotes the smallest singular value of the matrix A. As discussed extensively in [86], the function $z \mapsto (zI - A)^{-1}$ is called the *resolvent* of the matrix A. Thus the pseudospectra of A are just upper level sets of the *resolvent norm* function $n_A : \mathbb{C}\backslash\Lambda(A) \to \mathbb{R}_+$ defined by

$$n_A(z) := \left\|(zI - A)^{-1}\right\| = \frac{1}{\underline{\sigma}(A - zI)}.$$

Aside from the fact that pseudospectra is robust against numerical inaccuracies in measurements and implementation, pseudospectra may be more informative than eigenvalues in applications where matrices are non-normal [86, 44].

The continuity of the spectrum is well-known [49]. One immediate question is whether continuity extends to Λ_ϵ. Since Λ_ϵ is a set-valued map, we ask whether we have continuity in the Hausdorff metric, and it is known that the answer is yes [57, Theorem 2.3.7].

Does the pseudospectrum mapping Λ_ϵ have stronger continuity properties? One of my main results is finding conditions under which the map Λ_ϵ is Lipschitz continuous. For a given matrix A, we call points z that are smooth or nonsmooth critical points of the norm of the resolvent n_A *resolvent-critical*. We prove that Λ_ϵ is Lipschitz continuous around if and only if there are no resolvent-critical points $z \in \mathbb{C}$ such that $n_A(z) = \epsilon$.

As an application of the Lipschitz continuity of $\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$, we find conditions for the Lipschitz continuity (in the single-valued sense) and strict differentiability

of the pseudospectral abscissa $\alpha_\epsilon : M^n \to \mathbb{R}$ and the pseudospectral radius $\rho_\epsilon : M^n \to \mathbb{R}_+$, defined by

$$\begin{aligned} \alpha_\epsilon(A) &:= \max\{\mathrm{Re}(\lambda) \mid \lambda \in \Lambda_\epsilon(A)\}, \\ \rho_\epsilon(A) &:= \max\{|\lambda| \mid \lambda \in \Lambda_\epsilon(A)\}. \end{aligned} \tag{1.1.1}$$

We derive a variety of other properties of resolvent-critical points, proving in particular that points where pseudospectral components coalesce as ϵ grows are resolvent-critical.

1.2 Lipschitz behavior of the robust regularization

Motivated by the work on pseudospectra, we study the general process from which pseudospectra are obtained from eigenvalues, which motivates the "robust regularization" defined in [61].

Definition 1.1. For $\epsilon > 0$ and $F : X \to \mathbb{R}^m$, where $X \subset \mathbb{R}^n$, the *set-valued robust regularization* $F_\epsilon : X \rightrightarrows \mathbb{R}^m$ is defined as

$$F_\epsilon(x) := \{F(x+e) \mid |e| \le \epsilon, x+e \in X\}.$$

For the particular case of a real-valued function $f : X \to \mathbb{R}$, we define the *robust regularization* $\bar{f}_\epsilon : X \to \mathbb{R}$ of f by

$$\begin{aligned} \bar{f}_\epsilon(x) &:= \sup\{y \in f_\epsilon(x)\} \\ &= \sup\{f(x') \mid |x' - x| \le \epsilon\}. \end{aligned}$$

The operation of transforming a real-valued function to its robust regularization may be viewed as a kind of "deconvolution": see [47].

The issues of robust optimization, particularly in the case of linear and quadratic programming, are documented in [10]. Even if an optimal solution is found, implementing the solution precisely in a concrete model may be impossible (the design of digital filters being a typical example [43]). The minimizer of the robust regularization protects against small perturbations better, and might be a better solution to implement. We illustrate with the example

$$f(x) = \begin{cases} -x & \text{if } x < 0 \\ \sqrt{x} & \text{if } x \geq 0. \end{cases}$$

The robust regularization can be quickly calculated to be

$$\bar{f}_\epsilon(x) = \begin{cases} \epsilon - x & \text{if } x < \alpha(\epsilon) \\ \sqrt{\epsilon + x} & \text{if } x \geq \alpha(\epsilon), \end{cases}$$

where $\alpha(\epsilon) = \frac{1+2\epsilon-\sqrt{1+8\epsilon}}{2} > -\epsilon$. The minimizer of f is $\alpha(0)$, and f is not Lipschitz there. To see this, observe that $\frac{f(\delta)-f(0)}{\delta-0} \to \infty$ as $\delta \to 0$. But the robust regularization $\bar{f}_\epsilon$ is Lipschitz at its minimizer $\alpha(\epsilon)$; its left and right derivatives there are -1 and $\frac{1}{2\sqrt{\epsilon+\alpha(\epsilon)}}$, which are both finite.

The sensitivity of f at 0 can be attributed to the lack of Lipschitz continuity there. Lipschitz continuity is important in variational analysis, and is well studied in the recent books [80, 72]. The existence of a finite Lipschitz constant on f close to the optimizer can be important for sensitivity analysis for the problems from which the optimization problem was derived.

Several interesting examples of robust regularization are tractable to compute and optimize. For example, the robust regularization of any strictly convex quadratic is a semidefinite-representable function, tractable via semidefinite programming [61]. The spectral abscissa and spectral radius of a nonsymmetric square

matrix are the largest real part and the largest norm respectively of the eigenvalues of a matrix, and their robust regularizations are the pseudospectral abscissa and the pseudospectral radius as defined in (1.1.1). The pseudospectral abscissa is important in the study of the system $\frac{d}{dt}u(t) = Au(t)$, and is easily calculated using the algorithm in [21, 23], while the pseudospectral radius is important in the study of the system $u_{t+1} = Au_t$, and is easily calculated using the algorithm in [71].

We show that the robust regularization has a regularizing property: Even if the original function f is not Lipschitz at a point x, the robust regularization can be Lipschitz there under various conditions. We prove this regularizing property holds when the set of points at which f is not Lipschitz is isolated, or when f is a continuous semi-algebraic function. In the latter case, we prove that the Lipschitz modulus of $\bar{f}_\epsilon$ at $\bar{x}$ is of order $o\left(\frac{1}{\epsilon}\right)$. This estimate of the Lipschitz modulus can be helpful for robust design.

As an application, the pseudospectral abscissa and radius, being semi-algebraic functions, are Lipschitz at a fixed matrix for all small $\epsilon > 0$ with Lipschitz constant behaving like $o\left(\frac{1}{\epsilon}\right)$.

We also highlight the relation between calmness and Lipschitz continuity of single-valued mappings, which is important in our analysis. While this relation is studied in some generality for set-valued mappings (for example, in [65, Theorem 2.1], [79, Theorem 1.5]) in the study of metric regularity and subregularity [41], it seems to be used less in single-valued mappings.

1.3 Continuity of set-valued maps revisited in the light of tame geometry

Continuity properties of set-valued maps are crucial in many applications. It is often necessary to assume that the relevant set-valued maps in problems of controllability and feasibility satisfy some continuity properties. Another example of the importance of the continuity of set-valued maps is the definition of Clarke regularity of a set. Clarke regularity is a well-studied property in variational analysis, and is defined by the outer semicontinuity of the mapping to the normal cones of the set. A typical set-valued map arising from some construction or variational problem will not be continuous, but one would often expect a kind of semicontinuity to hold.

It is known that under added conditions, a closed-valued set-valued map that is either inner or outer semicontinuous is generically continuous. We shall recall this result recorded in [80, Theorem 5.55] and [7, Theorem 1.4.13], and attributed to [60, 30, 83]. We illustrate the sharpness and limitations of this result by appropriate examples. As a by-product of our analysis, we also mention an interesting consequence of these results by establishing a Sard-type result for the image of local minima for scalar-valued and vector-valued functions.

The main result is to establish that every semi-algebraic (or more generally, definable) closed-valued set-valued map is generically continuous. In the semi-algebraic context, genericity implies that possible failures can only arise in a set of lower dimension, and is equivalent to the measure-theoretical notion of *almost-everywhere*.

1.4 Level set methods for finding critical points of mountain pass type

Computing mountain passes is an important problem in computational chemistry and in the study of nonlinear partial differential equations. We begin with the following definition.

Definition 1.2. Let X be a topological space, and consider $a, b \in X$. For a function $f : X \to \mathbb{R}$, define a *mountain pass* $p^* \in \Gamma(a, b)$ to be a minimizer of the problem

$$\inf_{p \in \Gamma(a,b)} \sup_{0 \le t \le 1} f \circ p(t).$$

Here, $\Gamma(a, b)$ is the set of continuous paths $p : [0, 1] \to X$ such that $p(0) = a$ and $p(1) = b$.

An important problem in computational chemistry is to find the lowest energy to transition between two stable states. If a and b represent two states and f maps the states to their potential energies, then the mountain pass problem calculates this lowest energy. Early work on computing transition states includes Sinclair and Fletcher [85], and recent work is reviewed by Henkelman, Jóhannesson and Jónsson [46]. We refer to this paper for further references in the Computational Chemistry literature.

Perhaps more importantly, the mountain pass idea is also a useful tool in the analysis of nonlinear partial differential equations. For a Banach space X, variational problems are problems (P) such that there exists a smooth functional $J : X \to \mathbb{R}$ whose critical points (points where $\nabla J = 0$) are solutions of (P). Many partial differential equations are variational problems, and critical points of

J are "weak" solutions. In the landmark paper by Ambrosetti and Rabinowitz [4], the mountain pass theorem gives a sufficient condition for the existence of critical points in infinite dimensional spaces. If an optimal path to solve the mountain pass problem exists and the maximum along the path is greater than $\max(f(a), f(b))$, then the maximizer on the path is a critical point distinct from a and b. The mountain pass theorem and its variants are the primary ways to establish the existence of critical points and to find critical points numerically. For more on the mountain pass theorem and some of its generalizations, we refer the reader to [55].

In [29], Choi and McKenna proposed a numerical algorithm for the mountain pass problem by using an idea from Aubin and Ekeland [6] to solve a semilinear partial differential equation. This is extended to find solutions of *Morse index* 2 (that is, the maximum dimension of the subspace of X on which J'' is negative definite) in Ding, Costa and Chen [40], and then to higher Morse index by Li and Zhou [66].

Li and Zhou [67], and Yao and Zhou [93] proved convergence results to show that their minimax method is sound for obtaining weak solutions to nonlinear partial differential equations. Moré and Munson [73] proposed an "elastic string method", and proved that the sequence of paths created by the elastic string method contains a limit point that is a critical point.

The prevailing methods for numerically solving the mountain pass problem are motivated by finding a sequence of paths (by discretization or otherwise) such that the maximum along these paths decrease to the optimal value. Indeed, many methods in [46] approximate a mountain pass in this manner. As far as we are aware, only [9, 48] deviate from this strategy. We make use of a different approach by looking at the path connected components of the lower level sets of f instead.

One easily sees that l is a lower bound of the mountain pass problem if and only if a and b lie in two different path connected components of $\mathrm{lev}_{\leq l} f$. A strategy to find an optimal mountain pass is to start with a lower bound l and keep increasing l until the path connected components of $\mathrm{lev}_{\leq l} f$ containing a and b respectively coalesce at some point. However, this strategy requires one to determine whether the points a and b lie in the same path connected component, which is not easy. We turn to finding saddle points of mountain pass type, as defined below.

Definition 1.3. For a function $f : X \to \mathbb{R}$, a *saddle point of mountain pass type* $\bar{x} \in X$ is a point such that there exists an open set U such that $\bar{x}$ lies in the closure of two path components of $(\mathrm{lev}_{<f(\bar{x})} f) \cap U$.

We shall refer to saddle points of mountain pass type simply as saddle points. As an example, for the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = x_1^2 - x_2^2$, the point $\mathbf{0}$ is a saddle point of mountain pass type: We can choose $U = \mathbb{R}^2$, $a = (0, 1)$, $b = (0, -1)$. When f is $\mathcal{C}^1$, it is clear that saddle points are critical points. As we shall see later (in Propositions 6.20 and 6.21), saddle points of mountain pass type can, under reasonable conditions, be characterized as maximal points on mountain passes, acting as "bottlenecks" between two components. In fact, if f is $\mathcal{C}^2$, the Hessians are nonsingular and several mild assumptions hold, these bottlenecks are exactly critical points of Morse index 1. We refer the reader to the lecture notes by Ambrosetti [3]. Some of the methods in [46] actually find saddle points instead of solving the mountain pass problem.

We propose numerical methods to find saddle points using the strategy suggested in Definition 1.3. We start with a lower bound l and keep increasing l until the components of the level set $\mathrm{lev}_{\leq l} f \cap U$ containing a and b respectively coalesce, reaching the objective of the mountain pass problem. The first method we propose

in Algorithm 6.1 is purely metric in nature. One appealing property of this method is that calculations are now localized near the critical point and we keep track of only two points instead of an entire path. Our algorithm enjoys a monotonicity property: The distance between two components decreases monotonically as the algorithm progresses, giving an indication of how close we are to the saddle point.

It follows from the definitions that our algorithm, if it converges, converges to a saddle point. We then prove that any saddle point is deformationally critical in the sense of metric critical point theory [36, 59, 54], and is Morse critical under additional conditions. This implies in particular that any saddle point is Clarke critical in the sense of nonsmooth critical point theory [28, 84] based on nonsmooth analysis in the spirit of [16, 31, 72, 80]. It seems that there are few existing numerical methods for finding either critical points in a metric space or nonsmooth critical points. Currently, we are only aware of [92].

One of the main contributions of Chapter 6 is to give a second method (in Section 6.2) which converges locally superlinearly to a nondegenerate smooth critical point, i.e., critical points where the Hessian is nonsingular, in $\mathbb{R}^n$. Our numerical example in Section 6.7 illustrates this.

Our initial interest in the mountain pass problem came from computing the 2-norm distance of a matrix A to the closest matrix with repeated eigenvalues. This is also known as the Wilkinson problem, and this value is the smallest 2-norm perturbation that will make the eigenvalues of matrix A behave in a non-Lipschitz manner. Alam and Bora [1] showed how the Wilkinson's problem can be reduced to a global mountain pass problem. We do not solve the global mountain pass problem associated with the Wilkinson problem, but we demonstrate that locally our algorithm converges quickly to a smooth critical point of mountain pass type.

CHAPTER 2

PRELIMINARIES

In Section 2.1, we recall the definition and basic properties of variational analysis that we will use in Chapters 3 to 5. In Section 2.2, we give a technical discussion of semi-algebraic geometry.

2.1 Variational analysis

In this section, we give a brief introduction on the variational analytic tools used throughout the thesis, focusing in particular on the tools of set-valued analysis. We begin with the basic idea of convergence of sets.

Definition 2.1. [80, Definition 4.1] For a sequence $\{C^\nu\}_{\nu=1}^\infty$ of subsets of $\mathbb{R}^n$, the *outer limit* is the set

$$\begin{aligned}\limsup_{\nu\to\infty} C^\nu &:= \left\{x \mid \exists \text{ subsequence } N, x^\nu \in C^\nu \text{ with } x^\nu \underset{N}{\to} x\right\} \\ &= \{x \mid \forall \text{ open } V \ni x, \exists \text{ subsequence } N, \forall \nu \in N : C^\nu \cap V \neq \emptyset\},\end{aligned}$$

while the *inner limit* is the set

$$\begin{aligned}\liminf_{\nu\to\infty} C^\nu &:= \{x \mid \exists M > 0, \exists x^\nu \in C^\nu \, (\nu > M) \text{ with } x^\nu \to x\} \\ &= \{x \mid \forall \text{ open } V \ni x, \exists M > 0, \forall \nu > M : C^\nu \cap V \neq \emptyset\}.\end{aligned}$$

Here, the *limit* of the sequence exists if the outer and inner limit sets are equal:

$$\lim_{\nu\to\infty} C^\nu := \limsup_{\nu\to\infty} C^\nu = \liminf_{\nu\to\infty} C^\nu.$$

The Pompieu-Hausdorff distance (or Hausdorff metric) is a metric on compact subsets of $\mathbb{R}^n$. We recall its definition.

Definition 2.2. ([80, Example 4.13]) For $C, D \subset \mathbb{R}^n$ closed and nonempty, the *Pompieu-Hausdorff distance* (or *Hausdorff metric*) $\mathbf{d}(C, D)$ is defined as

$$\mathbf{d}(C, D) := \inf\{\eta \geq 0 \mid C \subset D + \eta\mathbb{B}, D \subset C + \eta\mathbb{B}\}.$$

For a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the outer and inner limits at $x \in \mathbb{R}^n$ are defined by the outer and inner limits above as follows:

$$\begin{aligned}
\limsup_{x\to\bar{x}} S(x) &:= \bigcup_{x^\nu\to\bar{x}} \limsup_{\nu\to\infty} S(x^\nu) \\
&= \{u \mid \exists x^\nu \to \bar{x}, \exists u^\nu \to u \text{ with } u^\nu \in S(x^\nu)\}, \\
\liminf_{x\to\bar{x}} S(x) &:= \bigcap_{x^\nu\to\bar{x}} \liminf_{\nu\to\infty} S(x^\nu) \\
&= \{u \mid \forall x^\nu \to \bar{x}, \exists M > 0 \text{ s.t. } u^\nu \to u \text{ and } u^\nu \in S(x^\nu)\,(\nu > M)\}.
\end{aligned}$$

These limits allow us to state the definition of continuity of set-valued maps.

Definition 2.3. [80, Definition 5.4] A set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *outer semicontinuous* (osc) at $\bar{x}$ if

$$\limsup_{x\to\bar{x}} S(x) \subset S(\bar{x}),$$

or equivalently $\limsup_{x\to\bar{x}} S(x) = S(\bar{x})$, and *inner semicontinuous* (isc) at $\bar{x}$ if

$$\liminf_{x\to\bar{x}} S(x) \supset S(\bar{x}),$$

or equivalently when S is closed-valued, $\liminf_{x\to\bar{x}} S(x) = S(\bar{x})$. It is called *continuous* at $\bar{x}$ if both conditions hold, i.e., if $S(x) \to S(\bar{x})$ as $x \to \bar{x}$.

If these terms are invoked relative to X, a subset of $\mathbb{R}^n$ containing $\bar{x}$, then the properties hold in restriction to convergence $x \to \bar{x}$ with $x \in X$ (in which case the sequences $x^\nu \to \bar{x}$ in the limit formulations are required to lie in X).

Lipschitz continuity of a set-valued map is defined as follows.

Definition 2.4. ([80, Definitions 9.26, 9.28]) A mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *Lipschitz continuous* if it is nonempty-closed-valued and there exists $\kappa \in \mathbb{R}_+$, a Lipschitz constant, such that $\mathbf{d}(S(x), S(x')) \leq \kappa |x - x'|$ for all $x, x' \in \mathbb{R}^n$, or

$$S(x') \subset S(x) + \kappa |x' - x| \mathbb{B} \text{ for all } x, x' \in \mathbb{R}^n.$$

The infimum of all κ such that there exists a neighbourhood V of $\bar{x}$ such that

$$S(x') \subset S(x) + \kappa |x' - x| \mathbb{B} \text{ for all } x, x' \in V$$

is the *Lipschitz modulus* for S at $\bar{x}$ and is denoted by $\operatorname{lip} S(\bar{x})$.

Lipschitz continuity of a set-valued map is sometimes too crude a measure of set-valued maps. The Aubin property, which is a localized Lipschitz property in the range of the set-valued map, is a more precise tool, and is defined as follows.

Definition 2.5. ([80, Definition 9.36]) (Aubin Property and graphical modulus) A mapping $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ has the *Aubin property at $\bar{x}$ for $\bar{u}$*, where $\bar{x} \in \mathbb{R}^n$ and $\bar{u} \in S(\bar{x})$, if gph S is locally closed at $(\bar{x}, \bar{u})$ and there are neighbourhoods V of $\bar{x}$ and W of $\bar{u}$, and a constant $\kappa \in \mathbb{R}_+$ such that

$$S(x') \cap W \subset S(x) + \kappa |x' - x| \mathbb{B} \text{ for all } x, x' \in V.$$

The *graphical modulus* of S at $\bar{x}$ for $\bar{u}$, denoted by $\operatorname{lip} S(\bar{x} \mid \bar{u})$, is the infimum of all such κ that satisfy the formula above.

Lipschitz continuity and the Aubin property are related by the following result. In the following statement, a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *locally bounded* at $\bar{x}$ if there is a neighborhood U of $\bar{x}$ such that $S(u)$ is bounded.

Proposition 2.6. *[72, Theorem 1.42] If $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is outer semicontinuous, and S is locally bounded at $\bar{x}$, then*

$$\operatorname{lip} S(\bar{x}) = \max_{y \in S(\bar{x})} \operatorname{lip} S(\bar{x} \mid y).$$

The generalization of the adjoint of a linear operator for set-valued maps is derived from the normal cones of its graph. We now define different types of normal cones and the concept of Clarke regularity of sets. In view of Chapter 6, we shall define normal cones of a set in a Hilbert space. We take the definition from [80, Definition 6.3] and [16].

Definition 2.7. (Normal cones) For a closed set D in a Hilbert space X and a point $\bar{z} \in D$, we recall that the *Hadamard normal cone (or Fréchet normal cone)* $\hat{N}_D(\bar{z})$ and the *limiting normal cone* $N_D(\bar{z})$ are defined by

$$\begin{aligned} \hat{N}_D(\bar{z}) &:= \{v \mid \langle v, z - \bar{z}\rangle \leq o(|z - \bar{z}|) \text{ for } z \in D\} \\ N_D(\bar{z}) &:= \{v \mid \exists \{(z_i, v_i)\}_i \subset \operatorname{gph} \hat{N}_D,\ v_i \xrightarrow{w} v \text{ and } z_i \to \bar{z}\} \end{aligned} \tag{2.1.1}$$

Here, "$\xrightarrow{w}$" stands for weak convergence. In the case where $X = \mathbb{R}^n$, weak convergence is the usual norm convergence, and $\hat{N}_D(\bar{z}) = \limsup_{z \to \bar{z}, z \in D} \hat{N}_D(z)$.

Definition 2.8. A set $C \subset \mathbb{R}^n$ is *Clarke regular* at one of its points $\bar{z}$ if it is locally closed at $\bar{z}$ and every normal vector to C at $\bar{z}$ is a regular normal vector, i.e., $N_C(\bar{z}) = \hat{N}_C(\bar{z})$.

We now define the coderivative of set-valued maps.

Definition 2.9. (Coderivatives) [80, Definition 8.33] For $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $(\bar{x}, \bar{y}) \in \operatorname{gph}(F)$, the *limiting coderivative* $D^*F(\bar{x} \mid \bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is defined by

$$D^*F(\bar{x} \mid \bar{y})(y^*) = \{x^* \mid (x^*, -y^*) \in N_{\operatorname{gph}(F)}(\bar{x}, \bar{y})\}.$$

It is clear from the definitions that the coderivative is a positively homogeneous map, which can be measured with the outer norm below.

Definition 2.10. [80, Section 9D] The *outer norm* $|\cdot|^+$ of a positively homogeneous map $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is defined by

$$\begin{aligned} |H|^+ &:= \sup_{w\in\mathbb{B}} \sup_{z\in H(w)} |z| \\ &= \sup\left\{ \frac{|z|}{|w|} \mid (w,z) \in \operatorname{gph}(H) \right\}. \end{aligned}$$

The following theorem, known as the Mordukhovich criterion [80, Theorem 9.40], shows how the coderivatives and the Aubin property of a set-valued map are related.

Theorem 2.11. *(Mordukhovich criterion) Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued map whose graph $\operatorname{gph}(S)$ is locally closed at $(\bar{x}, \bar{u}) \in \operatorname{gph}(S)$. Then S has the Aubin property at $\bar{x}$ with respect to $\bar{u}$ if and only if $D^*S(\bar{x} \mid \bar{u})(\mathbf{0}) = \{\mathbf{0}\}$ or equivalently $|D^*S(\bar{x} \mid \bar{u})|^+ < \infty$. In this case, $\operatorname{lip} S(\bar{x} \mid \bar{u}) = |D^*S(\bar{x} \mid \bar{u})|^+$.*

We will also use the subdifferential for functions mapping to a real line and the accompanying definition of subdifferential regularity. In view of Chapter 6, we shall define subdifferential of a functional on a Hilbert space (from [80, Definition 8.3] and [16, Definitions 3.1.1 and 5.2.20]).

Definition 2.12. Consider a function $f : X \to \mathbb{R} \cup \{\infty\}$, where X is a Hilbert space, and a point $\bar{x}$ with $f(\bar{x})$ finite. For a vector $v \in X$, one says that

(a) v is a *regular subgradient* (also known as a *Fréchet subgradient*) of f at $\bar{x}$, written $v \in \hat{\partial} f(\bar{x})$, if

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|);$$

(b) v is a *(general) subgradient* of f at $\bar{x}$, written $v \in \partial f(\bar{x})$, if there are sequences $x^\nu \to \bar{x}$ and $v^\nu \xrightarrow{w} v$ such that $f(x^\nu) \to f(\bar{x})$ and $v^\nu \in \hat{\partial} f(x^\nu)$.

(c) v is a horizon subgradient of f at $\bar{x}$, written $v \in \partial^\infty f(\bar{x})$, if there are sequences $x^\nu \to \bar{x}$, $t^\nu \searrow 0$ and $t^\nu v^\nu \xrightarrow{w} v$ such that $f(x^\nu) \to f(\bar{x})$ and $v^\nu \in \hat{\partial} f(x^\nu)$.

Here, "$\xrightarrow{w}$" stands for weak convergence. In the case where $X = \mathbb{R}^n$, weak convergence is the usual norm convergence.

Definition 2.13. A $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is *(subdifferentially) regular* at $\bar{x}$ if $f(\bar{x})$ is finite and the epigraph epi$f := \{(x, \alpha) \mid \alpha \geq f(x)\} \subset \mathbb{R}^n \times (\mathbb{R} \cup \{\infty\})$ is Clarke regular at $(\bar{x}, f(\bar{x}))$.

In the case where f is Lipschitz continuous at $\bar{x}$, we can use [80, Corollary 8.11, Theorem 9.13 and Theorem 8.6] to deduce that f is subdifferentially regular there if and only if $\hat{\partial} f(\bar{x}) = \partial f(\bar{x})$.

When a function is Lipschitz, another useful subdifferential is the Clarke subdifferential.

Definition 2.14. If $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz at $\bar{x}$, then the Clarke subdifferential ∂_C is defined by $\partial_C f(\bar{x}) = \text{conv}\partial f(\bar{x})$. It is well known that $\partial_C f(\bar{x}) = \text{conv}\, \limsup_{x \to \bar{x}} \nabla f(x)$, where the outer limit limsup is taken only over where the gradient $\nabla f(x)$ is defined.

The Clarke subdifferential can be defined in a Hilbert space for a lower semi-continuous function that is not necessarily Lipschitz, or even continuous. We shall use this formulation in Chapter 6.

Definition 2.15. [16, Theorem 5.2.19] Let X be a Hilbert space and let $f : X \to \mathbb{R}$ be a lsc function. The *Clarke subdifferential* of f at $\bar{x}$ is

$$\partial_C f(\bar{x}) := \text{cl conv}\{\text{w} - \lim_{i \to \infty} x_i^* \mid x_i^* \in \hat{\partial} f(x_i), (x_i, f(x_i)) \to (\bar{x}, f(\bar{x}))\} + \partial_C^\infty f(\bar{x}),$$

where the *Clarke horizon subdifferential* of f at $\bar{x}$ is a cone defined by

$$\partial_C^\infty f(\bar{x}) := \text{cl conv}\{\text{w}-\lim_{i\to\infty} \lambda_i x_i^* \mid x_i^* \in \hat{\partial} f(x_i), (x_i, f(x_i)) \to (\bar{x}, f(\bar{x})), \lambda_i \searrow 0\}.$$

Finally, we say that a point $\bar{x}$ is *Clarke-critical* if $\mathbf{0} \in \partial_C f(\bar{x})$. In the $\mathcal{C}^1$ case, Clarke-critical points are exactly where the derivatives are equal to zero, which coincides with the usual definition of a critical point of a smooth function mapping to the real line.

2.2 Semi-algebraic geometry

In this section we recall basic notions from semi-algebraic and o-minimal geometry. Let us define the notion of a semi-algebraic set [11, 34]. Let $\mathbb{R}[x_1, \ldots, x_n]$ be the ring of real polynomials of n variables.

Definition 2.16. [semi-algebraic set] A subset A of $\mathbb{R}^n$ is called *semi-algebraic* if it has the form

$$A = \bigcup_{i=1}^{k} \{x \in \mathbb{R}^n : p_i(x) = 0, q_{i1}(x) > 0, \ldots, q_{i\ell}(x) > 0\},$$

where $p_i, q_{ij} \in \mathbb{R}[x_1, \ldots, x_n]$ for all $i \in \{1, \ldots, k\}$ and $j \in \{1, \ldots, \ell\}$.

In other words, a set is semi-algebraic if it is a finite union of sets that are defined by means of a finite number of polynomial equalities and inequalities. A set-valued function $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *semi-algebraic*, if its graph $\text{gph}\,(S)$ is a semi-algebraic subset of $\mathbb{R}^n \times \mathbb{R}^m$.

We remark on the dimension of a semi-algebraic set. Before we begin, we recall the Whitney stratification ([42, Section 4.2], [33, Theorem 6.6]) is an important property of semi-algebraic sets.

Theorem 2.17. *($\mathcal{C}^k$ stratification) For any $k \in \mathbb{N}$ and any semi-algebraic subsets $X_1, \ldots, X_l$ of $\mathbb{R}^n$, we can write $\mathbb{R}^n$ as a disjoint union of finitely many semi-algebraic $\mathcal{C}^k$ manifolds $\{\mathcal{M}_i\}_i$ (that is, $\mathbb{R}^n = \dot{\cup}_{i=1}^I \mathcal{M}_i$) so that each X_j is a finite union of some of the $\mathcal{M}_i$'s. Moreover, the induced stratification $\{\mathcal{M}_i^j\}_i$ of X_j has the Whitney property that is, for any sequence $\{x_\nu\}_\nu \subset \mathcal{M}_i^j$ converging to $x \in \mathcal{M}_{i_0}^j$ we have $\limsup_{v \to \infty} N_{\mathcal{M}_i^j}(x_\nu) \subset N_{\mathcal{M}_{i_0}^j}(x)$.*

The Whitney stratification implies that every semi-algebraic set can be written as a finite disjoint union of manifolds ("strata") that fit together in a regular way ("Whitney stratification"). The Whitney property is also called *normal regularity* of the stratification. See [52, Definition 5]. The *dimension* $\dim(x)$ of a semi-algebraic set X can thus be defined as the dimension of the manifold of highest dimension of its stratification. This dimension is well defined and independent of the stratification of X [33, Section 3.3].

As a matter of the fact, semi-algebraic sets constitute an *o-minimal structure*. Let us recall the definitions of the latter (see for instance [34, 42]).

Definition 2.18. [o-minimal structure] An o-minimal structure on $(\mathbb{R}, +, .)$ is a sequence of Boolean algebras $\mathcal{O} = \{\mathcal{O}_n\}$, where each algebra $\mathcal{O}_n$ consists of subsets of $\mathbb{R}^n$, called *definable* (in $\mathcal{O}$), and such that for every dimension $n \in \mathbb{N}$ the following properties hold.

(i) For any set A belonging to $\mathcal{O}_n$, both $A \times \mathbb{R}$ and $\mathbb{R} \times A$ belong to $\mathcal{O}_{n+1}$.

(ii) If $\Pi : \mathbb{R}^{n+1} \to \mathbb{R}^n$ denotes the canonical projection, then for any set A belonging to $\mathcal{O}_{n+1}$, the set $\Pi(A)$ belongs to $\mathcal{O}_n$.

(iii) $\mathcal{O}_n$ contains every set of the form $\{x \in \mathbb{R}^n : p(x) = 0\}$, for polynomials $p : \mathbb{R}^n \to \mathbb{R}$.

(iv) The elements of $\mathcal{O}_1$ are exactly the finite unions of intervals and points.

When $\mathcal{O}$ is a given o-minimal structure, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ (or a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$) is called *definable* (in $\mathcal{O}$) if its graph is definable as a subset of $\mathbb{R}^n \times \mathbb{R}^m$.

It is obvious from the definition that semi-algebraic sets are stable under Boolean operations. As a consequence of the Tarski-Seidenberg principle, they are also stable under projections, thus they satisfy the above properties. Nonetheless, broader o-minimal structures also exist. In particular, the Gabrielov theorem implies that "globally subanalytic" sets are o-minimal. These two structures in particular provide rich practical tools, because checking semi-algebraicity or subanalyticity of sets in concrete problems of variational analysis is often easy. We refer to [13], [14], and [53] for details. Let us mention that Theorem 2.17 still holds in an arbitrary o-minimal structure (it is sufficient to replace the word "semi-algebraic" by "definable" in the statement). As a matter of the fact, the statement of Theorem 2.17 can be reinforced for definable sets (namely, the stratification can be taken analytic), but this is not necessary for our purposes.

We will also use other properties of semi-algebraic sets, for example, the local conic homeomorphism in 3.7, the simplicial decomposition theorem in Section 4.4, stratification for functions in Section 4.4, and the equivalence of genericity and the measure theoretic notion of "almost everywhere" in Chapter 5. Since we use these results once throughout the thesis, we shall recall them only as needed.

CHAPTER 3

VARIATIONAL ANALYSIS OF PSEUDOSPECTRA

The main result in this chapter is to prove conditions for the Lipschitz continuity of the pseudospectrum. Our proof (of the main results Theorem 3.26 and Proposition 3.29) may be described loosely by Figure 3.1. The reader may find the schematic outline helpful as the argument proceeds.

Let $\underline{\sigma} : M^n \to \mathbb{R}_+$ denote the map to the smallest singular value of a matrix. We write $MSV : M^n \rightrightarrows \mathbb{C}^n \times \mathbb{C}^n$, with

$$\begin{aligned} MSV(A) \;&:=\; \{(u,v) \mid u,v \text{ minimal left and} \\ &\qquad\qquad \text{right singular vectors of } A\}. \end{aligned}$$

In the above definition of MSV, u, v are *minimal left and right singular vectors* of A if they are unit vectors satisfying

$$\begin{aligned} \underline{\sigma}(A)u &= Av \\ \text{and } \underline{\sigma}(A)v &= A^H u, \end{aligned}$$

where A^H is the Hermitian transpose of A. A key tool in our analysis is the set

$$Y(A) \;:=\; \left\{v^H u \mid (u,v) \in MSV(A)\right\}.$$

We prove that the set $Y(A - zI)$ is the subgradient set at z of the function $-\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_-$, where $\underline{\sigma}_A(z) = \underline{\sigma}(A - zI)$.

In Figure 3.1, the six properties on the right on A and z are equivalent. For a given matrix A, we call points z not satisfying these equivalent properties "resolvent-critical" because they are smooth or nonsmooth critical points of the norm of the resolvent n_A. When the multiplicity of the smallest singular value

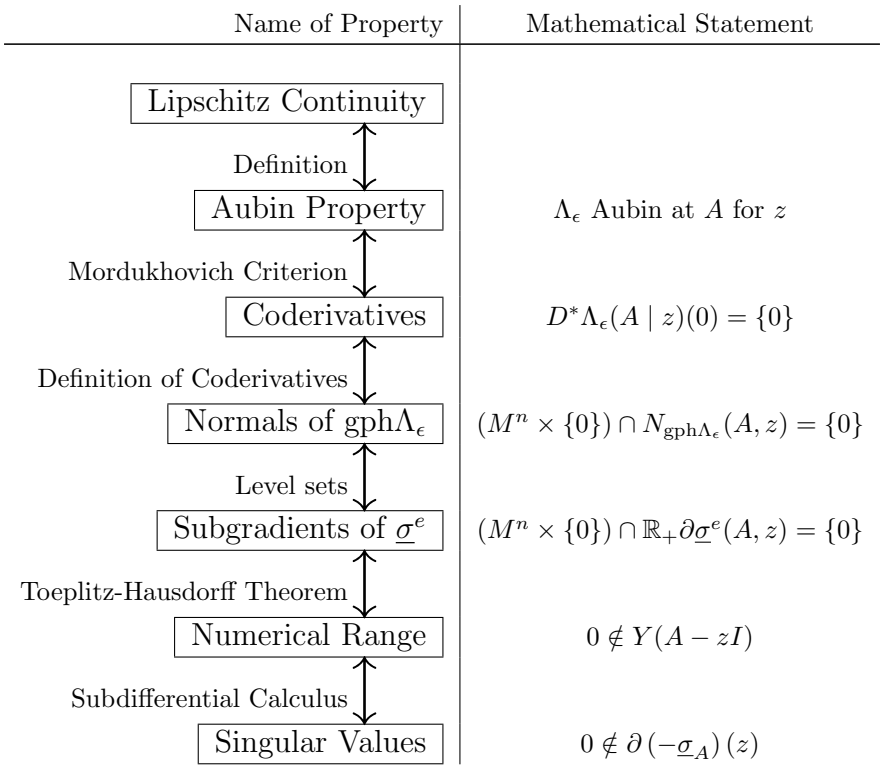


Figure 3.1: Equivalences of properties summarized in Theorem 3.26.

of $A - zI$ is one, this property is equivalent to z being a "degenerate point" (in the sense of [22, Definition 4.5, corrigendum]) or not "regular" in the sense of [23, Definition 4.4]. Points not resolvent-critical are exceptional for several aspects of pseudospectra, notably the quadratic convergence of the pseudospectral abscissa algorithm in [23].

Related to Λ_ϵ is the mapping, $\Lambda_\epsilon^c : M^n \rightrightarrows \mathbb{C}$, defined by $\Lambda_\epsilon^c(A) = \{z \mid \underline{\sigma}(A - zI) \geq \epsilon\}$. This mapping turns out to be easier to analyze because $-\underline{\sigma}(\cdot)$ has the property of subdifferential regularity (as defined in [80]) except at where it is zero. We show that the normal cone $N_{\Lambda_\epsilon^c(A)}(\bar{z})$ is $\mathbb{R}_+(Y(A - \bar{z}I))$. This establishes a link between the variational properties of $-\underline{\sigma}_A$ and Λ_ϵ^c, and the Aubin property.

Notation. For future reference, Tables 3.1 summarizes the mappings that

Table 3.1: Summary of definitions

Name/ Domain/ Range	Definition
$\bar{\sigma} : M^n \to \mathbb{R}_+$	$\bar{\sigma}(A)$ is maximum singular value of A
$\underline{\sigma} : M^n \to \mathbb{R}_+$	$\underline{\sigma}(A)$ is minimum singular value of A
$\underline{\sigma}^e : M^n \times \mathbb{C} \to \mathbb{R}_+$	$\underline{\sigma}^e(A, z) = \underline{\sigma}(A - zI)$
$\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_+$	$\underline{\sigma}_A(z) = \underline{\sigma}(A - zI)$
$\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$	$\Lambda_\epsilon(A) = \{z \mid \underline{\sigma}(A - zI) \leq \epsilon\}$
$\Lambda : M^n \rightrightarrows \mathbb{C}$	$\Lambda(A) = \Lambda_0(A) = \{\text{eigenvalues of } A\}$
$\Lambda_\epsilon^c : M^n \rightrightarrows \mathbb{C}$	$\Lambda_\epsilon^c(A) = \{z \mid \underline{\sigma}(A - zI) \geq \epsilon\}$
$\alpha_\epsilon : M^n \to \mathbb{R}$	$\alpha_\epsilon(A) = \max_{z \in \Lambda_\epsilon(A)} \operatorname{Re} z$
$\rho_\epsilon : M^n \to \mathbb{R}_+$	$\rho_\epsilon(A) = \max_{z \in \Lambda_\epsilon(A)} \lvert z \rvert$
$W : M^n \rightrightarrows \mathbb{C}$	Numerical range/ field of values[50, Definition 1.1.1]
$MSV : M^n \rightrightarrows \mathbb{C}^n \times \mathbb{C}^n$	See Definition 3.9
$Y : M^n \rightrightarrows \mathbb{C}$	See Definition 3.9

appear throughout this chapter.

Unless otherwise stated, our notation follows [80]. See also the table of notation in [80, page 725]. The term "regular" means subdifferentially regular in the sense Definition 2.13. Table 3.2 summarizes the symbols we use.

The "^{H}" in A^H and v^H represent the Hermitian transpose of a matrix or vector, while the "*" in L^* represents the adjoint of the linear operator L. Note that D^* stands for the coderivative instead. The real inner product on $A, B \in M^n$ is defined by Re tr $(A^H B)$.

Outline. This chapter is organized as follows. Section 3.1 studies the continu-

Table 3.2: Summary of definitions

Symbol	Explanation	Reference from [80]
pos	positive hull	Section 3G
$\text{lev}_{\leq\alpha} f$	Level sets: $\{x \mid f(x) \leq \alpha\}$	Section 1B
conv	convex hull	Section 1E
bdry	boundary of a set	
$\mathbb{B}$	unit ball $\{x \mid \|x\| \leq 1\}$	

ity properties of the pseudospectra Λ_ϵ and its "complement" Λ_ϵ^c via more general feasible-set mappings. In Sections 3.2, 3.3 and 3.4, we prove the main result that Λ_ϵ has the Aubin property at A for z if and only if $0 \notin Y(A - zI)$, with Section 3.2 containing general results on variational analysis and the singular value decomposition, Section 3.3 performing subdifferential calculus and Section 3.4 finishing the proof of the main result.

In Section 3.5, we show how the Lipschitz constant for the map Λ_ϵ can be calculated. Section 3.6 gives conditions for the Lipschitz continuity and strict differentiability of the pseudospectral abscissa α_ϵ and the pseudospectral radius ρ_ϵ. Finally, we present properties of resolvent-critical points in Section 3.7. We prove in particular that the points at which components of $\Lambda_\epsilon(A)$ coalesce as ϵ grows are resolvent-critical, and pose some questions about resolvent-critical points.

3.1 Feasible-set mappings and continuity of pseudospectra

The pseudospectral mapping $\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$ has two inputs: $\epsilon \in \mathbb{R}_+$ and the matrix in the argument of $\Lambda_\epsilon(\cdot)$. As $\mathbb{R}_+$ is one-dimensional, variation of $\Lambda_\epsilon(A)$ for a fixed matrix A and variable ϵ is easier to visualize, as implemented in EigTool [91].

Some attractive results in this direction have been obtained in [25, 27, 58, 1, 57] and elsewhere. By contrast, in this work we study how Λ_ϵ behaves for a fixed ϵ and a varying matrix argument, primarily taking a more abstract and systematic approach than [24].

We study pseudospectra using the language of set-valued analysis as described in the monograph [80]. In the next two propositions, let $f : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^m$ be a continuous function and let $T : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ be the mapping defined by

$$T(w) = \{x \mid f(x, w) \in D\}, \tag{3.1.1}$$

where D is a closed set.

Proposition 3.1. *T is outer semicontinuous.*

Proof. We just need to check that T has closed graph (by [80, Theorem 5.7]), which is easy. □

Note that the ϵ-pseudospectrum can be written as

$$\begin{aligned} \Lambda_\epsilon(A) &= \{z \mid \underline{\sigma}^e(A, z) \leq \epsilon\} \\ &= \{z \mid \underline{\sigma}^e(A, z) \in (-\infty, \epsilon]\}. \end{aligned}$$

If we apply Proposition 3.1, we can deduce that Λ_ϵ is outer semicontinuous. In a similar manner, Λ^c_ϵ, defined by $\Lambda^c_\epsilon(A) = \{z \mid \underline{\sigma}^e(A, z) \geq \epsilon\}$, is also outer semicontinuous.

Turning to inner semicontinuity, we begin with a technical result.

Proposition 3.2. *Let*

$$Q := \mathrm{cl}\{x \mid f(x, \bar{w}) \in \mathrm{int}(D)\},$$

so $Q \subset T(\bar{w})$. We have:

(a) $Q \subset \liminf_{w \to \bar{w}} T(w) \subset T(\bar{w})$

In the case where $m = 1$:

(b) *If $D = (-\infty, \alpha]$, then*

$$\begin{aligned} Q &= \{x \mid f(x, \bar{w}) = \alpha \text{ , } x \text{ is not a local minimizer of } f(\cdot, \bar{w})\} \\ &\quad \cup \{x \mid f(x, \bar{w}) < \alpha\}. \end{aligned}$$

(c) *If $D = [\alpha, \infty)$, then*

$$\begin{aligned} Q &= \{x \mid f(x, \bar{w}) = \alpha \text{ , } x \text{ is not a local maximizer of } f(\cdot, \bar{w})\} \\ &\quad \cup \{x \mid f(x, \bar{w}) > \alpha\}. \end{aligned}$$

(d) *If $\alpha > 0$, f is positively homogeneous (that is $\lambda f(\cdot) = f(\lambda \cdot)$ for $\lambda > 0$) and either $D = (-\infty, \alpha]$ or $D = [\alpha, \infty)$, then $Q = \liminf_{w \to \bar{w}} T(w)$.*

Proof. Property (a) is easy and standard. See for example the techniques in [8, 51].

Statements (b) and (c) are clear by the definition of Q, so we proceed to prove statement (d) for the case $D = (-\infty, \alpha]$. (The case $D = [\alpha, \infty)$ is similar and is omitted.) From statement (a), we just need to prove that if $\bar{x} \notin Q$, then $\bar{x} \notin \liminf_{w \to \bar{w}} T(w)$. Suppose that $\bar{x} \notin Q$. We need to consider only $\bar{x} \in T(\bar{w})$, so we can assume that $\bar{x}$ is a minimizer of $f(\cdot, \bar{w})$ and $f(\bar{x}, \bar{w}) = \alpha$. Then there is a neighbourhood $\mathbb{B}_\delta(\bar{x})$ about $\bar{x}$ such that $f(x, \bar{w}) \geq f(\bar{x}, \bar{w}) = \alpha$ if $x \in \mathbb{B}_\delta(\bar{x})$. If $\|x - \bar{x}\| < \delta/2$, then

$$\left\| \frac{1}{1 + \frac{1}{j}} x - \bar{x} \right\| \quad < \quad \delta \text{ if } j \text{ is large.}$$

This means that

$$\begin{aligned} f\left(x, \left(1+\frac{1}{j}\right)\bar{w}\right) &= \left(1+\frac{1}{j}\right) f\left(\frac{1}{1+\frac{1}{j}}x, \bar{w}\right) \\ &\geq \left(1+\frac{1}{j}\right)\alpha \quad \left(\text{because } \left\|\left(\frac{1}{1+\frac{1}{j}}\right)x - \bar{x}\right\| < \delta\right) \\ &> \alpha, \end{aligned}$$

which implies that $\mathbb{B}_{\delta/2}(\bar{x}) \cap T\left(\left(1+\frac{1}{j}\right)\bar{w}\right) = \emptyset$ if j is large enough. So for the sequence $\left(1+\frac{1}{j}\right)\bar{w} \to \bar{w}$ as $j \to \infty$, we cannot find a subsequence x_j such that $x_j \in T\left(\left(1+\frac{1}{j}\right)\bar{w}\right)$ and $x_j \to \bar{x}$, and this means that $\bar{x} \notin \liminf_{w\to\bar{w}} T(w)$. □

The following corollary is immediate from the definition of inner semicontinuity:

Corollary 3.3. *If $T(\bar{w}) = Q$, then T is continuous at $\bar{w}$. Furthermore, if f is positively homogeneous, then the converse holds as well.*

Proof. The mapping T is continuous if and only if it is both inner and outer semicontinuous. Apply the last two propositions. □

Now that we have established conditions for outer and inner semicontinuity for feasible set mappings, we shall study the continuity of the pseudospectrum Λ_ϵ and Λ_ϵ^c. Let us consider the case $\epsilon = 0$ first. The map $\Lambda_0^c : M^n \rightrightarrows \mathbb{C}$ is not interesting as $\Lambda_0^c(A) = \mathbb{C}$ for all matrices A. We are then led to consider the spectrum $\Lambda_0 = \Lambda$, which is well known to be continuous [49, Appendix D].

To extend to $\epsilon > 0$, we may apply Propositions 3.1 and 3.2, combined with the fact that $\underline{\sigma}_A(\cdot)$ has no local minimum other than at the eigenvalues [86, Theorem 2.4(i)], to prove the following result. This result is not new, and can be found, for example, in [57, Corollary 2.3.8].

Proposition 3.4. $\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$ *is continuous for* $\epsilon \geq 0$.

For $\Lambda_\epsilon^c : M^n \rightrightarrows \mathbb{C}$, we obtain the following using Proposition 3.2(d).

Proposition 3.5. $\Lambda_\epsilon^c : M^n \rightrightarrows \mathbb{C}$ *is outer semicontinuous, but it is inner semicontinuous at a matrix* A *if and only if there is no local maximizer* $\bar{z}$ *to* $\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_+$ *with* $\underline{\sigma}_A(\bar{z}) = \epsilon$.

Example 3.6. The mapping Λ_ϵ^c is not continuous at some points. For a concrete example of noncontinuity of Λ_ϵ^c, consider the point $0 \in \Lambda_1^c(\bar{A})$, where $\bar{A} = \operatorname{diag}(1, -1, i, -i)$ and $\epsilon = 1$. Here $\Lambda_1(\bar{A})$ consists of the union of balls of radius 1 around the diagonal entries, and so we observe that 0 is a local maximum of $\underline{\sigma}_{\bar{A}}$. This exhibits an example of discontinuity of Λ_1^c as $\liminf_{A \to \bar{A}} \Lambda_1^c(A) \subsetneq \Lambda_1^c(\bar{A})$.

Next, we consider Lipschitz continuity. If the function f in the feasible set mapping in formula (3.1.1) in page 26 is smooth, we understand the Aubin Property quite well through [80, Example 9.51]. If $D = (-\infty, \bar{\alpha}]$, we can also analyze the nonsmooth case.

Assumptions (a), (b) and (c) in the result below are standard for computing normals to level sets (see for example [80, Proposition 10.3].) Assumption (d) is needed to apply a chain rule

Theorem 3.7. *Consider the set-valued map* $C : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ *defined via a level set representation*

$$C(p) = \{x \mid F(x, p) \leq \bar{\alpha}\}$$

with $F : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}$. *Suppose*

(a) $F(\bar{x}, \bar{p}) = \bar{\alpha}$,

(b) $(0,0) \notin \partial F(\bar{x},\bar{p})$,

(c) *F is regular at* $(\bar{x},\bar{p})$,

(d) $(0,y_2) \in \partial^\infty F(\bar{x},\bar{p}) \implies y_2 = 0$.

Then C has the Aubin property at $\bar{p}$ for $\bar{x}$ if and only if $0 \notin \partial F_{\bar{p}}(\bar{x})$, *where* $F_{\bar{p}} : \mathbb{R}^n \to \mathbb{R}$ *is defined by* $F_{\bar{p}}(x) := F(x,\bar{p})$. *In this case,*

$$\operatorname{lip} C(\bar{p} \mid \bar{x}) = \max_{\substack{(a,b)\in N_{\operatorname{gph} C}(\bar{p},\bar{x}) \\ \|b\|=1}} \|a\|$$

If $F(\bar{x},\bar{p}) < \bar{\alpha}$, then C has the Aubin property at $\bar{p}$ for $\bar{x}$ with $\operatorname{lip} C(\bar{p} \mid \bar{x}) = 0$.

Proof. The Mordukhovich Criterion tells us that C has the Aubin property at $\bar{p}$ for $\bar{x}$ if and only if $D^*C(\bar{p} \mid \bar{x})(0) = \{0\}$. This holds if and only if

$$(z,0) \in N_{\operatorname{gph} C}(\bar{p},\bar{x}) \quad \text{implies} \quad z = 0. \tag{3.1.2}$$

This property is equivalent to

$$(0,z) \in N_{\operatorname{gph} C^{-1}}(\bar{x},\bar{p}) \quad \text{implies} \quad z = 0.$$

Conditions (a), (b) and (c) allow us to conclude that

$$N_{\operatorname{gph} C^{-1}}(\bar{x},\bar{p}) = (\operatorname{pos} \partial F(\bar{x},\bar{p})) \cup \partial^\infty F(\bar{x},\bar{p}) \tag{3.1.3}$$

through a result on level sets [80, Proposition 10.3], or

$$(0,z) \in (\operatorname{pos} \partial F(\bar{x},\bar{p})) \cup \partial^\infty F(\bar{x},\bar{p}) \text{ implies } z = 0$$

and by condition (d), this is in turn equivalent to

$$(0,z) \in \operatorname{pos} \partial F(\bar{x},\bar{p}) \text{ implies } z = 0 \tag{3.1.4}$$

We define $L_{\bar{p}} : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^d$ by $L_{\bar{p}}(x) = (x, \bar{p})$. The adjoint $L^*_{\bar{p}} : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^n$ is given by $L^*_{\bar{p}}(x, p) = x$. We have $F_{\bar{p}} = F \circ L_{\bar{p}}$, and so by a chain rule [80, Theorem 10.6] and condition (d), $\partial F_{\bar{p}}(\bar{x}) = L^*_{\bar{p}} \partial F(\bar{x}, \bar{p})$. Thus

$$\begin{aligned} \partial F_{\bar{p}}(\bar{x}) &= L^*_{\bar{p}} \partial F(\bar{x}, \bar{p}) \\ &= \{y \mid \exists z \text{ such that } (y, z) \in \partial F(\bar{x}, \bar{p})\}. \end{aligned}$$

If $0 \in \partial F_{\bar{p}}(\bar{x})$, then there exists z such that $(0, z) \in \partial F(\bar{x}, \bar{p})$, but condition (b) implies $z \neq 0$, which contradicts statement (3.1.4). If $0 \notin \partial F_{\bar{p}}(\bar{x})$, this means that there is no z such that $(0, z) \in \partial F(\bar{x}, \bar{p})$, and implies statement (3.1.4). So $0 \notin \partial F_{\bar{p}}(\bar{x})$ is equivalent to C not having the Aubin property at $\bar{p}$ for $\bar{x}$ as claimed.

The calculation of lip $C(\bar{p} \mid \bar{x})$ follows from the definition of the coderivative $D^*C(\bar{p} \mid \bar{x})$ and its relation to the normal cone through the Mordukhovich Criterion. If $F(\bar{x}, \bar{p}) < \bar{\alpha}$, then the normal cone is $\{(0, 0)\}$, giving us the required value of lip $C(\bar{p} \mid \bar{x})$. □

To obtain the Lipschitz modulus from the graphical modulus, one may use [80, Theorem 9.38], but Proposition 2.6 is sufficient for our purposes.

In Sections 3.2 to 3.5, we will be using the general principle illustrated in Theorem 3.7 to study where the pseudospectrum Λ_ϵ has the Aubin property, and also to illustrate how this can identify where Λ_ϵ is Lipschitz continuous and give a value of the Lipschitz constant.

One may immediately try to apply Theorem 3.7 to show that Λ_ϵ has the Aubin property for A at z. In this case, $p = A$, $x = z$, and so $C(p) = \Lambda_\epsilon(A)$, $F(x, p) = \underline{\sigma}(A - zI) = \underline{\sigma}^e(A, z)$. However, $\underline{\sigma}^e$ is not a regular function, but this can be overcome by studying $-\underline{\sigma}^e$ instead, which is regular if $A - zI$ is nonsingular. This is what we will do in the analysis that follows.

3.2 General results

First, we are interested in finding out whether the functions $-\underline{\sigma}^e$ and $\frac{1}{\underline{\sigma}^e}$ enjoy similar regularity properties so that we can deduce properties of $\underline{\sigma}^e$. We recall a result on the reciprocals of functions.

Proposition 3.8. *[72, Corollary 1.111(iii)] For any function $h : \mathbb{R}^n \to \mathbb{R}$ at z where $h(z) > 0$, we have $\partial h(z) = h(z)^2 \partial \left(-\frac{1}{h}\right)(z)$, and h is regular at z if and only if $-\frac{1}{h}$ is regular there.*

The set of minimal singular vectors of A, $MSV(A)$, is defined below.

Definition 3.9. For a matrix A, the left and right singular vectors corresponding to the smallest singular value of A are the pairs $(u, v) \in \mathbb{C}^n \times \mathbb{C}^n$, $\|u\| = \|v\| = 1$, which appear in the appropriate columns of U and V in some singular value decomposition $A = USV^H$ of A. We refer to u and v as *minimal singular vectors*, and we denote the set of pairs of minimal singular vectors of A as $MSV(A)$. Furthermore, define $Y : M^n \rightrightarrows \mathbb{C}$ by

$$Y(A) := \left\{ v^H u \mid (u, v) \in MSV(A) \right\}.$$

An equivalent definition given in the introduction is to have pairs of unit vectors (u, v) satisfying the equations $\underline{\sigma}(A)u = Av$ and $\underline{\sigma}(A)v = A^H u$.

The following result summarizes a complete characterization of left and right minimal singular vectors when we have one particular singular value decomposition, which is helpful for the case where the smallest singular value is multiple.

Proposition 3.10. *Consider a matrix $A \in M^n$ with singular value decomposition*

(for unit vectors u_j, v_j)

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^H = USV^H$$

where $\sigma_1 = \sigma_2 = \cdots = \sigma_m < \sigma_j$ *for all* $j > m$. *Define matrices* $\bar{U} = (u_1\, u_2 \cdots u_m)$ *and* $\bar{V} = (v_1\, v_2 \cdots v_m)$. *Then*

$$MSV(A) = \{(\bar{U}q, \bar{V}q) \mid q \in \mathbb{C}^m, \|q\| = 1\}$$

if A *is invertible (in other words,* $\sigma_1 > 0$*) and*

$$MSV(A) = \{(\bar{U}q_1, \bar{V}q_2) \mid q_1, q_2 \in \mathbb{C}^m, \|q_1\| = \|q_2\| = 1\}$$

if A *is singular.*

Proof. The equations $Av = \underline{\sigma}(A)u$ and $A^H u = \underline{\sigma}(A)v$ require u to be an eigenvector for AA^H and v to be an eigenvector for $A^H A$, and so they lie in the subspaces spanned by the columns of $\bar{U}$ and $\bar{V}$ respectively. We have assumed that these columns are placed at the left of U and V. Then let $v = \bar{V}q$. As we want a v of unit length, we must have $\|q\| = 1$. Since A is invertible, $\underline{\sigma} := \underline{\sigma}(A) > 0$, and so

$$u = \frac{1}{\underline{\sigma}} Av = \frac{1}{\underline{\sigma}} USV^H \bar{V}q = \frac{1}{\underline{\sigma}} US \begin{pmatrix} I \\ 0 \end{pmatrix} q = U \begin{pmatrix} I \\ 0 \end{pmatrix} q = U \begin{pmatrix} q \\ 0 \end{pmatrix} = \bar{U}q.$$

Thus $MSV(A) \subset \{(\bar{U}q, \bar{V}q) \mid q \in \mathbb{C}^m, \|q\| = 1\}$. The reverse is straightforward.

If A is singular, then as before, $u = \bar{U}q_1$ and $v = \bar{V}q_2$ for some unit vectors q_1, q_2. It is evident that u and v satisfy the relations $\underline{\sigma}(A)u = Av$ and $\underline{\sigma}(A)v = A^H u$, so we are done. □

The significance of $Y(A)$ will become clear later in sections 3.3 and 3.4. We first show a result on $Y(A)$.

Proposition 3.11. *If* A *is invertible, then* $Y(A)$ *is convex.*

Proof. We make the observation that the set $Y(A)$ can be determined as follows. Let $\bar{U}$ and $\bar{V}$ be as described in Proposition 3.10. The numerical range of a matrix $B \in M^n$ is the set $\left\{v^H B v \mid v \in \mathbb{C}^n, \|v\| = 1\right\}$, denoted by $W(B)$, and is convex by the Toeplitz-Hausdorff Theorem [50, Property 1.2.2]. Then

$$\begin{aligned} Y(A) &= \{v^H u \mid (u,v) \in MSV(A)\} \\ &= \left\{q^H \bar{V}^H \bar{U} q \mid \|q\| = 1\right\} \text{ (by Proposition 3.10)} \\ &= W\left(\bar{V}^H \bar{U}\right), \text{ the numerical range of } \bar{V}^H \bar{U}, \end{aligned}$$

establishing the convexity of $Y(A)$. □

For singular matrices A, $Y(A)$ need not be convex. Take for example the singular value decomposition:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

With this matrix,

$$\begin{aligned} Y(A) &= \left\{ q_1 \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} q_2 \mid q_1, q_2 \in \mathbb{C}, |q_1| = |q_2| = 1 \right\} \\ &= \{q \in \mathbb{C} \mid |q| = 1\}, \end{aligned}$$

which is not convex.

3.3 Subdifferential calculus

This section collects some results about subdifferential calculus involving $\underline{\sigma}^e : M^n \times \mathbb{C} \to \mathbb{R}_+$, where $\underline{\sigma}^e(A, z) = \underline{\sigma}(A - zI)$. As suggested in Figure 3.1, there is a link between the subdifferential $\partial \underline{\sigma}^e(A, z)$ and normal cone $N_{\operatorname{gph} \Lambda_\epsilon}(A, z)$ for

$\underline{\sigma}^e(A,z) = \epsilon$. Before we can apply the appropriate theorems in [80], we have to calculate $\partial\underline{\sigma}^e(A,z)$, establish regularity properties and characterize whether $0 \in \partial\underline{\sigma}^e(A,z)$.

When the smallest singular value is simple, $\underline{\sigma}$ and $\underline{\sigma}^e$ are analytic, as the next lemmas assert.

We remind the reader that the spaces M^n and $M^n \times \mathbb{C}$ have (real) inner products defined by

$$\langle A, B\rangle = \text{Re tr}(A^H B) \text{ for } A, B \in M^n$$

and

$$\langle (X,x),(Y,y)\rangle = \text{Re}\left(\text{tr}(X^H Y) + x^H y\right) \text{ for } X, Y \in M^n \text{ and } x, y \in \mathbb{C}.$$

Lemma 3.12. *If the invertible matrix A has a simple smallest singular value, then the function $\underline{\sigma} : M^n \to \mathbb{R}_+$ is real-analytic at A, with gradient*

$$\nabla\underline{\sigma}(A) = uv^H$$

for any $(u,v) \in MSV(A)$.

The proof for the above lemma is standard (for example, [22, Theorem 7.1]), while the lemma below follows by noticing that $\underline{\sigma}^e = \underline{\sigma} \circ L$ and applying the chain rule, where $L : M^n \times \mathbb{C} \to \mathbb{C}$ is defined by $L(A,z) = A - zI$.

Lemma 3.13. *If $z \notin \Lambda(A)$ and $A - zI$ has a simple smallest singular value, then the function $\underline{\sigma}^e : M^n \times \mathbb{C} \to \mathbb{R}_+$ is real-analytic at Z, with gradient*

$$\nabla\underline{\sigma}^e(A,z) = (uv^H, -v^H u)$$

for any $(u,v) \in MSV(A - zI)$.

The next two results are generalizations of Lemmas 3.12 and 3.13 to the nonsmooth case, and calculates the subgradients needed in the main result in Section 3.4.

Proposition 3.14. *Suppose $z \notin \Lambda(A)$. Then*

$$\partial(-\underline{\sigma}^e)(A,z) = \mathrm{conv}\{(-uv^H, v^Hu) \mid (u,v) \in MSV(A-zI)\}.$$

Furthermore, $-\underline{\sigma}^e$ is regular at (A,z) and globally Lipschitz.

Proof. We consider the functions

$$\bar{\sigma}^e : M^n \times \mathbb{C} \to \mathbb{R}_+,\ \iota : M^n \to M^n \text{ and } L : M^n \times \mathbb{C} \to M^n$$

defined by

$$\bar{\sigma}^e(A,z) = \bar{\sigma}((A-zI)^{-1}),\ \iota(B) = B^{-1} \text{ and } L(A,z) = A - zI.$$

That is, $\bar{\sigma}^e = \bar{\sigma} \circ \iota \circ L$. To evaluate the subdifferential of this function, we apply a chain rule [80, Theorem 10.6]. Given a matrix B, we seek to evaluate $\nabla(\iota \circ L)(A,z)^*(B)$, which is, by the chain rule, $\nabla L(A,z)^*(\nabla\iota(A-zI)^*(B))$.

As $\bar{\sigma}$ is everywhere Lipschitz, $\partial^\infty\bar{\sigma}(\iota \circ L(A,z)) = \{0\}$. Furthermore, since $\bar{\sigma}$ is convex, it is regular at $\iota \circ L(A,z)$, and so the conditions for [80, Theorem 10.6] are satisfied.

It is easy to check the identity $L^*(B) = (B, -\mathrm{tr}B)$. (Note that L is linear so $\nabla L = L$ and $\nabla L^* = L^*$.) Using the binomial expansion

$$(M+\Delta)^{-1} = M^{-1} - M^{-1}\Delta M^{-1} + o(\Delta),$$

it follows that $\nabla\iota(M)(B) = -M^{-1}BM^{-1}$, from which $\nabla\iota\,(M)^*\,(B) = -M^{-H}BM^{-H}$ follows easily.

Next, we evaluate $\partial\bar{\sigma}^e(A,z)$. Let the singular value decomposition of $(A-zI)$ be USV^H. Then the singular value decomposition of $(A-zI)^{-1}$ is $VS^{-1}U^H$, and $(A-zI)^{-H} = US^{-1}V^H$. So

$$\partial\bar{\sigma}^e(A,z) = \nabla L(A,z)^*\nabla\iota(A-zI)^*\partial\bar{\sigma}\left((A-zI)^{-1}\right).$$

We know that

$$\partial\bar{\sigma}(B) = \text{conv}\{uv^H \mid \|u\| = \|v\| = 1, Bv = \bar{\sigma}(B)u, B^Hu = \bar{\sigma}(B)v\}.$$

(See for example [87].) Therefore,

$$\partial\bar{\sigma}((A-zI)^{-1}) = \text{conv}\{vu^H \mid (u,v) \in MSV(A-zI)\}.$$

Then for any $(u,v) \in MSV(A-zI)$, we have:

$$\begin{aligned}
\nabla L(A,z)^*\nabla\iota(A-zI)^*(vu^H) &= \nabla L(A,z)^*(-US^{-1}V^Hvu^HUS^{-1}V^H) \\
&= \underline{\sigma}(A-zI)^{-2}\nabla L(A,z)^*(-uv^H) \\
&= \underline{\sigma}(A-zI)^{-2}(-uv^H, \text{tr }(uv^H)) \\
&= \underline{\sigma}(A-zI)^{-2}(-uv^H, v^Hu),
\end{aligned}$$

and so

$$\partial\bar{\sigma}^e(A,z) = \underline{\sigma}(A-zI)^{-2}\text{conv}\{(-uv^H, v^Hu) \mid (u,v) \in MSV(A-zI)\}.$$

By Proposition 3.8, we conclude

$$\begin{aligned}
\partial(-\underline{\sigma}^e)(A,z) &= \partial\left(-\frac{1}{\bar{\sigma}^e}\right)(A,z) \\
&= \bar{\sigma}^e(A,z)^{-2}\partial\bar{\sigma}^e(A,z) \\
&= \text{conv}\{(-uv^H, v^Hu) \mid (u,v) \in MSV(A-zI)\}.
\end{aligned}$$

The function $-\underline{\sigma}^e$ is regular at (A,z) because $\bar{\sigma}$ is regular and both the chain rule [80, Theorem 10.6] and Proposition 3.8 guarantee the preservation of regularity. Also, the function $-\underline{\sigma}^e$ is globally Lipschitz because $-\underline{\sigma}^e = -\underline{\sigma} \circ L$ is the composition of two globally Lipschitz functions. □

From the definition of $\Lambda_\epsilon(A) = \{z \mid \underline{\sigma}_A(z) \leq \epsilon\}$, where $\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_+$ is defined by $\underline{\sigma}_A(z) = \underline{\sigma}(A - zI)$, it is clear that the functions $\underline{\sigma}$ and $\underline{\sigma}_A$ figure prominently in the study of pseudospectra. The following two results can be seen as nonsmooth analogues of [22, Theorem 7.1 and Corollary 7.2]. Even though $\underline{\sigma}$ and $\underline{\sigma}_A$ are not necessarily smooth, we are able to prove that $-\underline{\sigma}$ and $-\underline{\sigma}_A$ are regular and calculate their subgradients.

Proposition 3.15. *The function $-\underline{\sigma}$ is regular at every nonsingular matrix $A \in M^n$ with*

$$\partial(-\underline{\sigma})(A) \;=\; -\mathrm{conv}\{uv^H \mid (u,v) \in MSV(A)\}$$

Proof. Define $L_{M^n} : M^n \to M^n \times \mathbb{C}$ by $L_{M^n}(A) = (A, 0)$, so we have $-\underline{\sigma}_A = (-\underline{\sigma}^e) \circ L_{M^n}$. Clearly L_{M^n} is smooth, with $\nabla L_{M^n} = I \times \mathbf{0}$ at all points. $(L_{M^n})^* : M^n \times \mathbb{C} \to M^n$ is just the natural projection. Thus, by appealing to [80, Theorem 10.6] and Proposition 3.14, we get what we need. □

Proposition 3.16. *For a matrix A, consider the function $\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_+$ defined by $\underline{\sigma}_A(z) = \underline{\sigma}(A - zI)$. If $z \notin \Lambda(A)$, then*

$$\partial(-\underline{\sigma}_A)(z) \;=\; Y(A - zI)$$

and $-\underline{\sigma}_A$ is regular at z and globally Lipschitz.

Proof. The proof is similar to the proof above, but we work through the details for completeness. We note $-\underline{\sigma}_A = (-\underline{\sigma}^e) \circ L_A$, where $L_A : \mathbb{C} \to M^n \times \mathbb{C}$, $L_A(z) = (A, z)$. Clearly L_A is smooth, with $\nabla L_A = \mathbf{0} \times I$ at all points. Furthermore, $(\nabla L_A)^* : M^n \times \mathbb{C} \to \mathbb{C}$ is just the natural projection. Thus, by appealing to a chain rule [80, Theorem 10.6] and Proposition 3.14, we have

$$\begin{aligned} \partial(-\underline{\sigma}_A)(z) &= (\nabla L_A)^* \partial(-\underline{\sigma}^e)(A, z) \\ &= Y(A - zI). \end{aligned}$$

As in Proposition 3.14, $\underline{\sigma}_A$ is globally Lipschitz because it is a composition of two globally Lipschitz functions. $\square$

We note that the assumptions that $A - zI$ is nonsingular in Proposition 3.14 and A is nonsingular in Proposition 3.15 cannot be dropped in the proposition below.

Proposition 3.17. *If $z \in \Lambda(A)$, then $-\underline{\sigma}^e$ is not regular at (A, z). Similarly, $-\underline{\sigma}$ is not regular at A if A is singular.*

Proof. Take $\bar{U}$ and $\bar{V}$ to the matrices corresponding to the minimal left and right singular vectors of $A - zI$ in the statement of Proposition 3.10. For small $\epsilon > 0$, we have

$$\begin{aligned} -\underline{\sigma}^e\left(A + \epsilon\bar{U}\bar{V}^H, z\right) &= -\underline{\sigma}^e(A, z) - \epsilon \\ \text{and } -\underline{\sigma}^e\left(A - \epsilon\bar{U}\bar{V}^H, z\right) &= -\underline{\sigma}^e(A, z) - \epsilon. \end{aligned}$$

Hence if $(B, x) \in \hat{\partial}(-\underline{\sigma}^e)(A, z)$, we have

$$\begin{aligned} -\underline{\sigma}^e\left(A \pm \epsilon\bar{U}\bar{V}^H, z\right) &\geq -\underline{\sigma}^e(A, z) + \left\langle (B, x), \left(\pm\epsilon\bar{U}\bar{V}^H, 0\right)\right\rangle + o(\epsilon) \\ \implies -\epsilon &\geq \epsilon\left\langle (B, x), \left(\pm\bar{U}\bar{V}^H, 0\right)\right\rangle + o(\epsilon). \end{aligned}$$

Dividing by ϵ throughout and taking limits as $\epsilon \downarrow 0$, we have

$$\begin{aligned} -1 &\geq \left\langle (B, x), \left(\pm\bar{U}\bar{V}^H, 0\right)\right\rangle \\ \implies -2 &\geq \left\langle (B, x), \left(\bar{U}\bar{V}^H, 0\right)\right\rangle + \left\langle (B, x), \left(-\bar{U}\bar{V}^H, 0\right)\right\rangle = 0, \end{aligned}$$

which is obviously a contradiction. This means that $\hat{\partial}\left(-\underline{\sigma}^e\right)(A, z) = \emptyset$. To show that $\partial\left(-\underline{\sigma}^e\right)(A, z) \neq \emptyset$, we note that for small $\epsilon > 0$, we have

$$(-u_1 v_1^H, v_1^H u_1) \in \hat{\partial}(-\underline{\sigma}^e)\left(A + \epsilon\bar{U}\bar{V}^H, z\right)$$

by Proposition 3.14, where the minimal left and right singular vectors u_1, v_1 are defined in the statement of Proposition 3.10. Taking $\epsilon \downarrow 0$, this ensures that $(-u_1 v_1^H, v_1^H u_1) \in \partial(-\underline{\sigma}^e)(A, z)$ and thus $\partial(-\underline{\sigma}^e)(A, z) \neq \emptyset$. Since $\partial(-\underline{\sigma}^e)$ and $\hat{\partial}(-\underline{\sigma}^e)$ differ and appealing to [80, Corollary 8.11], $-\underline{\sigma}^e$ is not regular at (A, z). The proof for $-\underline{\sigma}$ is similar. □

Proposition 3.18. *The resolvent norm $n_A : \mathbb{C} \to \mathbb{R}$ defined by $n_A(z) = \|(zI - A)^{-1}\|$ is regular at every point where $z \notin \Lambda(A)$, with*

$$\partial n_A(z) = n_A(z)^2 Y(A - zI).$$

Proof. From the identity $n_A = 1/\underline{\sigma}_A$ and Propositions 3.8 and 3.16, we note the following calculations:

$$\begin{aligned}
\partial n_A(z) =& n_A(z)^2 \partial \left(-\frac{1}{n_A} \right)(z) \\
=& n_A(z)^2 \partial(-\underline{\sigma}_A)(z) \\
=& n_A(z)^2 Y(A - zI).
\end{aligned}$$

□

This motivates the following definition.

Definition 3.19. A point $z \in \mathbb{C}$ is *resolvent-critical* for a square matrix A if either $z \in \Lambda(A)$ or $0 \in Y(A - zI)$.

Thus resolvent-critical points that are not eigenvalues are simply critical points of the resolvent norm n_A (in the nonsmooth sense). Since $\underline{\sigma}_A$ is globally Lipschitz, the following holds as well.

Theorem 3.20. *For a given matrix A, the following are equivalent:*

(1) *z is resolvent-critical.*

(2) *z is Clarke-critical for $-\underline{\sigma}_A$.*

(3) *z is Clarke-critical for $\underline{\sigma}_A$.*

Proof. Since $\underline{\sigma}_A$ is Lipschitz, we have $\partial^\circ(-\underline{\sigma}_A)(z) = -\partial^\circ\underline{\sigma}_A(z)$ by [31, Proposition 2.3.1]. This means that (2) and (3) are equivalent.

Next we prove that (1) implies (2). If z is resolvent-critical, then either z is an eigenvalue of A or $0 \in \partial\left(-\underline{\sigma}_A\right)(z)$. In the second case, z is Clarke-critical for $-\underline{\sigma}_A$ because $\partial(-\underline{\sigma}_A)(z) \subset \partial^\circ(-\underline{\sigma}_A)(z)$. In the first case, z is a maximizer of $-\underline{\sigma}_A$, and so z is Clarke-critical.

Lastly, we prove that (2) implies (1). If z is not resolvent-critical, then z is not an eigenvalue and $0 \notin \partial(-\underline{\sigma}_A)(z)$. But $\partial\left(-\underline{\sigma}_A\right)(z) = \partial^\circ\left(-\underline{\sigma}_A\right)(z)$ by the regularity of $-\underline{\sigma}_A$ at z so we are done. □

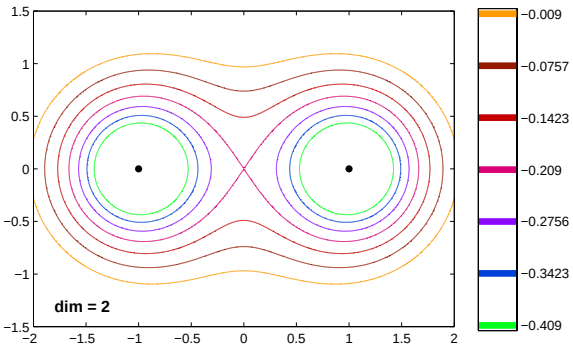
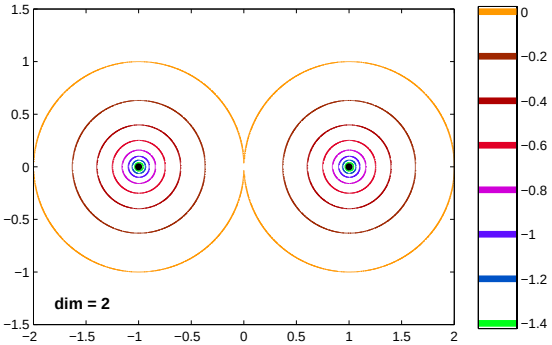
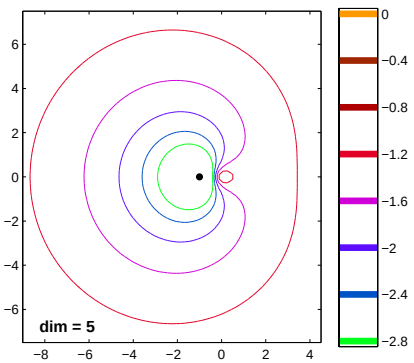
Example 3.21. Table 1 shows some examples where 0 is a resolvent-critical point of A. (In the third example, the resolvent-critical point is close to 0 but not exactly at 0.) These plots were obtained with EigTool [91]. The curves represent the boundaries of the pseudospectra $\Lambda_\epsilon(A)$ for $\epsilon = 10^\alpha$, where α is the number corresponding to the line generated by EigTool in the legend on the right. The third example is found in [37].

We also have an alternative proof to [22, Theorem 9.2] after the remark below.

The set

$$G(z) = \left\{v^H(A - zI)v \mid v \in V(z), \|v\| = 1\right\}$$

Table 3.3: Examples of Pseudospectra for Example 3.21.

A	Diagram
$\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$	Smooth Saddle 
$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	Nonsmooth Saddle 
$-\begin{pmatrix} 1 & 5 & 5^2 & 5^3 & 5^4 \\ 0 & 1 & 5 & 5^2 & 5^3 \\ 0 & 0 & 1 & 5 & 5^2 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	Local minimum of n_A 

where the subspace $V(z) \subset \mathbb{C}^n$ is spanned by all right singular vectors of $A - zI$ as defined in [22, Section 9] is equal to $\underline{\sigma}(A - zI)Y(A - zI)$.

Proposition 3.22. *If $\bar{z}$ is not resolvent-critical and $\underline{\sigma}_A(\bar{z}) = \epsilon$, then the set $\Lambda^c_\epsilon(A)$ is Clarke regular at $\bar{z}$, with normal cone $N_{\Lambda^c_\epsilon(A)}(\bar{z}) = \mathrm{pos}(Y(A - \bar{z}I))$.*

Proof. This involves applying Proposition 3.16 to a result on level sets [80, Proposition 10.3]. □

The conditions below on $\partial\underline{\sigma}^e(A, z)$ and $\partial(-\underline{\sigma}^e)(A, z)$ are needed in a manner similar to condition (b) in Theorem 3.7 in the proof of our main result.

Proposition 3.23. *The condition $(0, 0) \in \partial\underline{\sigma}^e(A, z)$ holds if and only if $z \in \Lambda(A)$. Also, if $z \notin \Lambda(A)$, then $(0, 0) \notin \partial(-\underline{\sigma}^e)(A, z)$.*

Proof. If $\underline{\sigma}^e(A, z) = 0$, then (A, z) is a local minimizer and thus $(0, 0) \in \partial\underline{\sigma}^e(A, z)$. On the other hand, if $\underline{\sigma}^e(A, z) > 0$, we need to prove that $(0, 0) \notin \partial\underline{\sigma}^e(A, z)$. We try to evaluate $\hat{\partial}\underline{\sigma}^e(A, z)$. From Proposition 3.14, we know that at points where the multiplicity of the singular value $\underline{\sigma}(A - zI)$ is greater than one, $\underline{\sigma}^e$ is not differentiable. By [80, Corollary 9.21], $\hat{\partial}\underline{\sigma}^e(A, z) = \emptyset$ at these points. For points where the multiplicity of the singular value is one, the norm calculation tells us that the only point in $\hat{\partial}\underline{\sigma}^e(A, z)$ has norm at least 1; the only element in $\hat{\partial}\underline{\sigma}^e(A, z)$ is of the form $\left(uv^H, -v^H u\right)$, and the matrix part already contributes 1 to the norm. So it is impossible that $(0, 0) \in \partial\underline{\sigma}^e(A, z)$.

Next, we move on to $\partial\left(-\underline{\sigma}^e\right)(A, z)$. Take $\bar{U}, \bar{V}$ to be the matrix corresponding to the left and right singular vectors of $A - zI$ in the sense of Proposition 3.10. Note that $\left(\bar{U}\bar{V}^H, 0\right)$ represents a direction of linear descent, as

$$-\underline{\sigma}^e\left(A + \epsilon\bar{U}\bar{V}^H, z\right) = -\underline{\sigma}^e(A, z) - \epsilon$$

for small ϵ, and so we have $(0,0) \notin \hat{\partial}(-\underline{\sigma}^e)(A,z)$. Due to regularity (Proposition 3.14), we have $(0,0) \notin \partial(-\underline{\sigma}^e)(A,z)$. $\square$

Despite the fact that $\underline{\sigma}^e$ is not regular, we are still able to calculate the subdifferential $\partial\underline{\sigma}^e(A,z)$

Proposition 3.24. *If $z \notin \Lambda(A)$, then*

$$\partial\underline{\sigma}^e(A,z) = \left\{ \left(uv^H, -v^H u\right) \mid (u,v) \in MSV(A - zI) \right\}$$

Proof. We observe that

$$\begin{aligned} \partial\underline{\sigma}^e(A,z) &\subset -\partial\left(-\underline{\sigma}^e\right)(A,z) \\ &= \operatorname{conv}\left\{ \left(uv^H, -v^H u\right) \mid (u,v) \in MSV(A - zI) \right\} \end{aligned}$$

by [80, Corollary 9.21] and Proposition 3.14. Next, note that if $(B,w) \in \partial\underline{\sigma}^e(A,z)$, then

$$(B,w) \in \operatorname{conv}\left\{ \left(uv^H, -v^H u\right) \mid (u,v) \in MSV(A - zI) \right\},$$

and so we may write $(B,w) = \sum_{i=1}^{k} \lambda_i \left(u_i v_i^H, -v_i^H u_i\right)$ for a convex combination of left and right singular vectors u_i, v_i corresponding to the smallest singular value. But since the 2-norm is a strictly convex norm, $\|B\| < 1$ if $k > 1$ and (u_i, v_i)'s are not complex multiples each other. We take a closer look: (B,w) can be written as a limit of $(B_i, w_i) = \nabla\underline{\sigma}^e(A_i, z_i)$ where $(A_i, z_i) \to (A,z)$ by [80, Corollary 9.21]. Since $\|B_i\| = 1$, it follows that $\|B\| = 1$.

With this, we conclude that $(B,w) = (uv^*, -v^*u)$ for some $(u,v) \in MSV(A - zI)$ and so

$$\partial\underline{\sigma}^e(A,z) \subset \left\{ \left(uv^H, -v^H u\right) \mid (u,v) \in MSV(A - zI) \right\}.$$

To prove the other containment, note that for any $(u, v) \in MSV(A - zI)$, we have

$$\begin{aligned}\hat{\partial}\underline{\sigma}^e(A - \delta uv^H, z) &= \left\{\nabla\underline{\sigma}^e(A - \delta uv^H, z)\right\} \\ &= \left\{\left(uv^H, -v^H u\right)\right\}\end{aligned}$$

for $0 < \delta < \epsilon$ by Lemma 3.13. Taking limits as $\delta \downarrow 0$, we have $\left(uv^H, -v^H u\right) \in \partial\underline{\sigma}^e(A, z)$, which completes the proof. □

3.4 Main result

Before proving our main result, we make a statement about the normal cones $N_{\mathrm{gph}\Lambda^c_\epsilon}(A, z)$ and $N_{\mathrm{gph}\Lambda_\epsilon}(A, z)$. We make use of properties that we have established in Section 3.3 to establish the link between level sets and normal vectors.

Proposition 3.25. *If $\epsilon = \underline{\sigma}^e(A, z) > 0$, then*

$$\begin{aligned}N_{\mathrm{gph}\Lambda^c_\epsilon}(A, z) &= \text{pos conv}\left\{\left(-uv^H, v^H u\right) \mid (u, v) \in MSV(A - zI)\right\}, \\ N_{\mathrm{gph}\Lambda_\epsilon}(A, z) &= \text{pos}\left\{\left(uv^H, -v^H u\right) \mid (u, v) \in MSV(A - zI)\right\}.\end{aligned}$$

Proof. Apply a result on level sets [80, Proposition 10.3], Proposition 3.23 and the fact that $-\underline{\sigma}^e$ is Lipschitz to get

$$N_{\mathrm{gph}\Lambda^c_\epsilon}(A, z) = \text{pos}\left(\partial(-\underline{\sigma}^e)(A, z)\right).$$

Next, apply Proposition 3.14 to deduce the first result.

By [80, Proposition 10.3] and Proposition 3.24 we have

$$\begin{aligned}N_{\mathrm{gph}\Lambda_\epsilon}(A, z) &\subset \text{pos}\ \partial\underline{\sigma}^e(A, z) \\ &= \text{pos}\left\{\left(uv^H, -v^H u\right) \mid (u, v) \in MSV(A - zI)\right\}.\end{aligned}$$

Furthermore, if $\underline{\sigma}(A - zI)$ is simple then $\underline{\sigma}^e$ is smooth and regular at (A, z) by Lemma 3.13, and so the above inclusion holds with equality.

For the opposite containment, take any $(u, v) \in MSV(A - zI)$. Consider the pair

$$(A_\delta, z_\delta) := \big((1+\delta)A - \epsilon\delta uv^H, (1+\delta)z\big) \text{ for small } \delta > 0.$$

At these points, $\underline{\sigma}^e$ is smooth (and thus regular) because the singular value is of multiplicity one with corresponding singular vectors (u, v), and $\underline{\sigma}^e(A_\delta, z_\delta) = \epsilon$. Thus

$$\big(uv^H, -v^H u\big) \in \hat{N}_{\mathrm{gph}\Lambda_\epsilon}((1+\delta)A - \epsilon\delta uv^H, (1+\delta)\, z).$$

Taking $\delta \downarrow 0$, we see that $\big(uv^H, -v^H u\big) \in N_{\mathrm{gph}\Lambda_\epsilon}(A, z)$. Since $N_{\mathrm{gph}\Lambda_\epsilon}(A, z)$ is a cone, we have the formula for $N_{\mathrm{gph}\Lambda_\epsilon}(A, z)$ as claimed. □

The following is the main result that summarizes the links between the diagram in the introduction.

Theorem 3.26. *Consider a point $z \notin \Lambda(A)$. Let $\epsilon = \underline{\sigma}^e(A, z)$. Then the following are equivalent:*

(1) *z is not resolvent-critical for A.*

(2) *Λ_ϵ^c has the Aubin property at A for z.*

(3) *Λ_ϵ has the Aubin property at A for z.*

Proof. For the purposes of the proof, we introduce several other properties:

(4) $(M^n \times \{0\}) \cap N_{\mathrm{gph}\Lambda_\epsilon^c}(A, z) = \{0\}$.

(5) $D^*\Lambda_\epsilon^c(A \mid z)(0) = \{0\}$.

(6) $(M^n \times \{0\}) \cap N_{\mathrm{gph}\Lambda_\epsilon}(A, z) = \{0\}$.

(7) $D^*\Lambda_\epsilon(A \mid z)(0) = \{0\}$.

Properties (4) and (5) are equivalent because $\alpha \in D^*\Lambda_\epsilon^c(A \mid z)\,(\beta)$ if and only if $(\alpha, -\beta) \in N_{\mathrm{gph}\Lambda_\epsilon^c}(A, z)$ by the definition of coderivatives. Properties (5) and (2) are equivalent by the Mordukhovich Criterion. The same goes for properties (6), (7) and (3).

Next, we show the equivalence of properties (1) and (4). We apply Proposition 3.25 to reduce property (4) to

$$(M^n \times \{0\}) \cap \operatorname{pos\,conv}\left\{\left(-uv^H, v^H u\right) \mid (u, v) \in MSV(A - zI)\right\} = \{0\}\,.$$

$(1 \Rightarrow 4)$ Suppose that z is not resolvent-critical, that is $0 \notin Y(A - zI)$, and yet property (4) fails. Then there is some nonzero pair with second coordinate (the one in $\mathbb{C}$) zero lying in

$$\operatorname{pos\,conv}\left\{\left(-uv^H, v^H u\right) \mid (u, v) \in MSV(A - zI)\right\}.$$

This means that there is a convex combination of pairs $\left(-uv^H, v^H u\right)$ such that their second coordinate is zero. Then $0 \in Y(A - zI)$ (appealing to Proposition 3.11), a contradiction.

$(1 \Leftarrow 4)$ If property (1) fails, there are minimal left and right singular vectors u, v such that $v^H u = 0$, and then $\left(-uv^H, v^H u\right)$ is a nonzero element in

$$(M^n \times \{0\}) \cap \operatorname{pos\,conv}\left\{\left(-uv^H, v^H u\right) \mid (u, v) \in MSV(A - zI)\right\}.$$

So we have proved the equivalence of properties (1) and (4). We proceed to prove the equivalence of properties (1) and (6). We lose regularity, but nevertheless, the proof still looks similar.

$(1 \Rightarrow 6)$ We prove $(4 \Rightarrow 6)$. If $0 \notin Y(A - zI)$, then $(M^n \times \{0\}) \cap N_{\mathrm{gph}\Lambda_\epsilon^c}(A, z) = \{0\}$. But Proposition 3.25 gives

$$\begin{aligned} \{0\} &\subset (M^n \times \{0\}) \cap N_{\mathrm{gph}\Lambda_\epsilon}(A, z) \\ &\subset (M^n \times \{0\}) \cap -N_{\mathrm{gph}\Lambda_\epsilon^c}(A, z) \\ &= \{0\}. \end{aligned}$$

$(1 \Leftarrow 6)$. If property (1) fails, there are minimal left and right singular vectors u, v such that $v^H u = 0$, and thus $\left(uv^H, -v^H u\right)$ is a nonzero element in $(M^n \times \{0\}) \cap N_{\mathrm{gph}\Lambda_\epsilon}(A, z)$. □

When we consider fixing the matrix A and increasing ϵ, it is natural to ask whether the map $\epsilon \mapsto \Lambda_\epsilon(A)$ is Lipschitz.

Proposition 3.27. *Given $z \in \mathbb{C}$, the map $\epsilon \mapsto \Lambda_\epsilon(A)$ has the Aubin property at $\underline{\sigma}_A(z)$ for z if and only if $0 \notin \partial \underline{\sigma}_A(z)$, whereas the map $\epsilon \mapsto \Lambda_\epsilon^c(A)$ has the Aubin property at $\underline{\sigma}_A(z)$ for z if and only if $0 \notin \partial(-\underline{\sigma}_A)(z)$ (or equivalently, assuming $z \notin \Lambda(A)$, z is not resolvent-critical for A).*

Proof. A straightforward application of [80, Theorem 9.41(b)] on $\underline{\sigma}_A$ gives us $0 \notin \partial \underline{\sigma}_A(z)$ if and only if the map $\epsilon \mapsto \mathrm{lev}_{\leq \epsilon} \underline{\sigma}_A = \Lambda_\epsilon(A)$ has the Aubin property at ϵ for z, which is the first part of what we seek to prove. The second part is similar, using Proposition 3.16. □

A particular example worked out in full detail exploiting this is highlighted in [24].

It is natural to ask whether there are any differences between Theorem 3.26 and the two parts of Proposition 3.27, and it comes down to comparing $\partial(-\underline{\sigma}_A)$

and $\partial\underline{\sigma}_A$. In general, if z is not an eigenvalue of A,

$$-\partial\underline{\sigma}_A(z) \subset \partial(-\underline{\sigma}_A)(z) = Y(A - zI)$$

by Proposition 3.16 and [80, Corollary 9.21], but the inclusion can be strict. Consider the matrix $\bar{A} = \operatorname{diag}(1, -1, i, -i)$ in Example 3.6. Here, $\partial(-\underline{\sigma}_A)(0) = \{a + bi \mid |a| + |b| \leq 1\}$ so 0 is resolvent-critical while $\partial\underline{\sigma}_A(0) = \{1, -1, i, -i\}$.

3.5 Lipschitz continuity of pseudospectra

The results in the last section study the Aubin property of the pseudospectra Λ_ϵ. The next natural step is to evaluate the graphical modulus and investigate the Lipschitz continuity of Λ_ϵ.

If $\underline{\sigma}(A - zI) = \epsilon > 0$, then from Proposition 3.25 and the definition of the coderivative, we can deduce the formula for $D^*\Lambda_\epsilon^c(A \mid z)(c)$. To keep the expressions compact, we understand that (u_i, v_i) ranges over $MSV(A - zI)$ whenever u_i, v_i appear in the formulas below. We have

$$\begin{aligned}
& D^*\Lambda_\epsilon^c(A \mid z)(c) \\
= \; & \left\{ -k \sum_i \lambda_i u_i v_i^H \mid c = -k \sum_i \lambda_i v_i^H u_i, \sum_i \lambda_i = 1, \lambda_i \geq 0, k \geq 0 \right\} \\
= \; & \begin{cases} \left\{ c \frac{\sum_i \lambda_i u_i v_i^H}{\sum_i \lambda_i v_i^H u_i} \mid \sum_i \lambda_i v_i^H u_i \neq 0 \right\} & \text{if } c \neq 0 \\ \operatorname{pos}\left\{ \sum_i \lambda_i u_i v_i^H \mid \sum_i \lambda_i v_i^H u_i = 0 \right\} & \text{if } c = 0 \end{cases}
\end{aligned}$$

and

$$\begin{aligned} & D^*\Lambda_\epsilon(A \mid z)(c) \\ = \; & \left\{ kuv^H \mid c = kv^Hu, k \geq 0, (u,v) \in MSV(A - zI) \right\} \\ = \; & \begin{cases} \left\{ c\frac{uv^H}{v^Hu} \mid (u,v) \in MSV(A-zI), v^Hu \neq 0 \right\} & \text{if } c \neq 0 \\ \operatorname{pos}\left\{ uv^H \mid (u,v) \in MSV(A-zI), v^Hu = 0 \right\} & \text{if } c = 0. \end{cases} \end{aligned}$$

We can then calculate the graphical moduli for Λ_ϵ and Λ_ϵ^c in the theorem below.

Theorem 3.28. *We have the following graphical moduli:*

$$\begin{aligned} \operatorname{lip}\, \Lambda_\epsilon(A \mid z) &= \begin{cases} 1/\mathrm{d}\,(0, Y(A-zI)) & \text{if } \underline{\sigma}(A-zI) = \epsilon \\ 0 & \text{if } \underline{\sigma}(A-zI) < \epsilon \end{cases} \\ \operatorname{lip}\, \Lambda_\epsilon^c(A \mid z) &= \begin{cases} 1/\mathrm{d}\,(0, Y(A-zI)) & \text{if } \underline{\sigma}(A-zI) = \epsilon \\ 0 & \text{if } \underline{\sigma}(A-zI) > \epsilon. \end{cases} \end{aligned}$$

(Here, we interpret $1/0 = +\infty$.)

Proof. It is clear that if $\underline{\sigma}(A - zI) < \epsilon$, then (A, z) lies in the interior of $\mathrm{gph}\Lambda_\epsilon$, so $N_{\mathrm{gph}\Lambda_\epsilon}(A, z) = \{(0,0)\}$, and so

$$\operatorname{lip}\, \Lambda_\epsilon(A \mid z) = |D^*\Lambda_\epsilon(A \mid z)|^+ = 0.$$

Similarly, $\operatorname{lip}\, \Lambda_\epsilon^c(A \mid z) = 0$ if $\underline{\sigma}(A - zI) > \epsilon$.

If $\underline{\sigma}(A - zI) = \epsilon$ and $0 \in Y(A - zI)$, then Λ_ϵ and Λ_ϵ^c do not have the Aubin property at A for z, and so

$$\operatorname{lip}\, \Lambda_\epsilon(A \mid z) = \operatorname{lip}\, \Lambda_\epsilon^c(A \mid z) = \infty.$$

By the Mordukhovich Criterion and the definition of outer norms, we have $\operatorname{lip}\, \Lambda_\epsilon^c(A \mid z)$ to be

$$\sup_{c \neq 0} \sup_{d \in D^*\Lambda_\epsilon^c(A|z)(c)} \frac{\|d\|}{|c|},$$

or in other words the infimum of all κ such that

$$d \in D^* \Lambda_\epsilon^c(A \mid z)(c) \implies \|d\| \leq \kappa |c| . \tag{3.5.1}$$

In view of the formula for $D^* \Lambda_\epsilon^c(A \mid z)$, formula (3.5.1) is equivalent to

$$\left\| \sum \lambda_i u_i v_i^H \right\| \leq \kappa \left| \sum \lambda_i v_i^H u_i \right| \tag{3.5.2}$$

for all $(u_i, v_i) \in MSV(A - zI), \lambda_i \geq 0, \sum \lambda_i = 1$. To prove that lip $\Lambda_\epsilon^c(A \mid z) = 1/d\,(0, Y(A - zI))$, it remains to prove that formula (3.5.2) is equivalent to

$$\kappa \geq 1/d\,(0, Y(A - zI)) . \tag{3.5.3}$$

Suppose that κ satisfies formula (3.5.2). Then for $y \in Y(A - zI)$, we have some $(u, v) \in MSV(A - zI)$ such that $y = v^H u$. Then

$$\begin{aligned} \kappa |y| =& \kappa \left| v^H u \right| \\ \geq & \left\| u v^H \right\| \\ =& 1. \end{aligned}$$

Formula (3.5.3) follows. Next, suppose that κ satisfies formula (3.5.3). If $(u_i, v_i) \in MSV(A - zI)$, $\lambda_i \geq 0$ and $\sum \lambda_i = 1$, we have $\sum \lambda_i v_i^H u_i \in Y(A - zI)$ by the convexity of $Y(A - zI)$. Thus

$$\begin{aligned} \left\| \sum \lambda_i u_i v_i^H \right\| \leq & \sum \lambda_i \left\| u_i v_i^H \right\| \\ =& 1 \\ \leq & \kappa \left| \sum \lambda_i v_i^H u_i \right| \end{aligned}$$

Formula (3.5.2) follows and so lip $\Lambda_\epsilon^c(A \mid z) = 1/d\,(0, Y(A - zI))$. Similar and simpler calculations give us lip $\Lambda_\epsilon(A \mid z) = 1/d\,(0, Y(A - zI))$. $\square$

We next turn to the Lipschitz constant for the pseudospectral mapping Λ_ϵ. We want to find $\mathrm{lip}_\infty \Lambda_\epsilon(\bar{A})$, the Lipschitz modulus of the pseudospectral map at $\bar{A}$. For a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, we are able to calculate $\mathrm{lip}\, S(\bar{x})$ from the graphical modulus easily with Proposition 2.6. Thus the Lipschitz constants for Λ_ϵ are easily obtained.

Proposition 3.29. *The following expressions are equal:*

$$\begin{aligned}
&\text{(i)} \quad \mathrm{lip}\, \Lambda_\epsilon(A) \\
&\text{(ii)} \quad \max_{z \in \Lambda_\epsilon(A)} \{\mathrm{lip}\, \Lambda_\epsilon(A \mid z)\} \\
&\text{(iii)} \quad \max_{z : \underline{\sigma}(A - zI) = \epsilon} \{1/d(0, Y(A - zI))\} \\
&\text{(iv)} \quad \max_{z} \left\{1/\left|v^H u\right| \mid (u, v) \in MSV(A - zI), \underline{\sigma}(A - zI) = \epsilon\right\}.
\end{aligned}$$

Proof. The expressions (i) and (ii) are equal by Proposition 2.6 and the fact that Λ_ϵ is compact and locally bounded. Then expression (ii) and (iii) are equal by Theorem 3.28, and expression (iv) is just an expansion of the definition of $Y(\cdot)$ applied to expression (iii). □

3.6 Pseudospectral abscissa and pseudospectral radius

In this section we apply our results on Lipschitz continuity of pseudospectra to re-explore earlier work on the pseudospectral abscissa and pseudospectral radius in [22, 23, 71, 86].

Definition 3.30. Define the *ϵ-pseudospectral abscissa* $\alpha_\epsilon : M^n \to \mathbb{R}$ by

$$\alpha_\epsilon(A) = \max_{z \in \Lambda_\epsilon(A)} \mathrm{Re}(z),$$

and the *ϵ-pseudospectral radius* $\rho_\epsilon : M^n \to \mathbb{R}_+$ by

$$\rho_\epsilon(A) = \max_{z \in \Lambda_\epsilon(A)} |z| .$$

Note that if $\epsilon > 0$, then $\rho_\epsilon(A) > 0$. We shall establish continuity properties of α_ϵ and ρ_ϵ. We begin with another routine piece of theory on parametric minimization.

Corollary 3.31. *(to [80, Corollary 10.14]) Suppose that $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is outer semicontinuous and maps to compact sets. Define $p : \mathbb{R}^m \to \mathbb{R}$ and $P : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ below by*

$$p(u) = \min_{x \in F(u)} g(x), \ P(u) = \arg\min_{x \in F(u)} g(x),$$

where the lower semicontinuous function $g : \mathbb{R}^n \to \mathbb{R}$ is differentiable at all points in $P(\bar{u})$ for some given $\bar{u} \in \mathbb{R}^m$. Then p is

(a) Lipschitz continuous around $\bar{u}$ if F has the Aubin property at $\bar{u}$ for $\bar{x}$ for all $\bar{x} \in P(\bar{u})$, with

$$\text{lip } p(\bar{u}) \ \leq \ \max\{|y| : y \in S\} < \infty$$

where $S = \{y \mid \bar{x} \in P(\bar{u}), y \in D^*F(\bar{u} \mid \bar{x})(\nabla g(\bar{x}))\}$;

(b) strictly differentiable at $\bar{u}$ with $\nabla p(\bar{u}) = \bar{y}$ if $S = \{\bar{y}\}$.

Proof. Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$ be defined by

$$f(x,u) = \delta_{\text{gph}F}(u,x) + g(x) = \begin{cases} g(x) & \text{if } x \in F(u) \\ \infty & \text{otherwise} \end{cases}$$

Then

$$p(u) = \inf_x f(x,u), \ \ P(u) = \arg\min_x f(x,u).$$

Since F is outer semicontinuous, gph F is closed so f is proper and lower semicontinuous.

Next, we prove f is level bounded in x locally uniformly in u. That is, for each $\bar{u} \in \mathbb{R}^m$ and $\alpha \in \mathbb{R}$, there is a neighbourhood V of $\bar{u}$ along with a bounded set $B \subset \mathbb{R}^n$ such that $\{x \mid f(x,u) \leq \alpha\} \subset B$ for all $u \in V$. Note that $f(x,u) \leq \alpha$ means that $x \in F(u)$ and $g(x) \leq \alpha$. Since F is outer semicontinuous, choose V such that $F(u) \subset F(\bar{u}) + \mathbb{B}$ for all $u \in V$, by the characterization of outer semicontinuity. The set B can be chosen to be $F(\bar{u}) + \mathbb{B}$ and we are done.

Following the notation in [80, Corollary 10.13], for any $\bar{x} \in P(\bar{u})$,

$$\begin{aligned} M(\bar{x},\bar{u}) &:= \{y \mid (0,y) \in \partial f(\bar{x},\bar{u})\} \\ &= \{y \mid (y,0) \in \partial \delta_{\mathrm{gph}F}(\bar{u},\bar{x}) + \{(0,\nabla g(\bar{x}))\}\} \\ &\qquad\qquad \text{(by [80, Exercise 8.8(c)])} \\ &= \{y \mid (y,-\nabla g(\bar{x})) \in N_{\mathrm{gph}F}(\bar{u},\bar{x})\} \\ &\qquad\qquad \text{(by [80, Exercise 8.14])} \\ &= D^*F(\bar{u} \mid \bar{x})(\nabla g(\bar{x})) \\ &\qquad\qquad \text{(by [80, Definition 8.33]).} \end{aligned}$$

Also,

$$\begin{aligned} M_\infty(\bar{x},\bar{u}) &:= \{y \mid (0,y) \in \partial^\infty f(\bar{x},\bar{u})\} \\ &= \{y \mid (y,0) \in \partial^\infty \delta_{\mathrm{gph}F}(\bar{u},\bar{x})\} \\ &= \{y \mid (y,0) \in N_{\mathrm{gph}F}(\bar{u},\bar{x})\} \\ &= D^*F(\bar{u} \mid \bar{x})(0). \end{aligned}$$

This means that $Y_\infty(\bar{u}) := \bigcup_{\bar{x} \in P(\bar{u})} M_\infty(\bar{x},\bar{u}) = \{0\}$, so part (a) of [80, Corollary 10.14] applies. Furthermore, $Y(\bar{u})$, where $Y(\cdot)$ is defined in [80, Corollary 10.13],

is

$$\begin{aligned} Y(\bar{u}) &:= \bigcup_{\bar{x}\in P(\bar{u})} M(\bar{x},\bar{u}) \\ &= \bigcup_{\bar{x}\in P(\bar{u})} D^*F(\bar{u}\mid\bar{x})\left(\nabla g(\bar{x})\right), \end{aligned}$$

and so,

$$\begin{aligned} \operatorname{lip} p(\bar{u}) &\leq \max_{y\in Y(\bar{u})} |y| \\ &= \max\left\{|y| : \bar{x}\in P(\bar{u}), y\in D^*F(\bar{u}\mid\bar{x})(\nabla g(\bar{x}))\right\} < \infty. \end{aligned}$$

The rest of the claim follows by [80, Corollary 10.14]. □

The continuity of α_ϵ and ρ_ϵ can be proved by the following proposition when the conditions for Lipschitz continuity are absent. The proof is routine.

Proposition 3.32. *Suppose that $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is continuous and maps to compact sets. If p, P and g are defined as in Corollary 3.31 with g continuous, then p is continuous and P is outer semicontinuous.*

As a consequence of Corollary 3.31, we obtain the following result.

Corollary 3.33. *The pseudospectral abscissa α_ϵ and pseudospectral radius ρ_ϵ are Lipschitz continuous at a matrix A if $\operatorname{lip}_\infty\Lambda_\epsilon(A) < \infty$, with Lipschitz constants bounded above by $\operatorname{lip}_\infty\Lambda_\epsilon(A)$.*

Proof. Following the notation in Corollary 3.31, take $F = \Lambda_\epsilon$ and $g(x) = \langle -1, x\rangle$. Then $\alpha_\epsilon = -p$ and we obtain

$$\begin{aligned} \operatorname{lip}\alpha_\epsilon(A) &\leq \max\{|y| : y\in D^*\Lambda_\epsilon(A\mid z)(-1) \\ &\qquad , z\in\Lambda_\epsilon(A), \operatorname{Re}(z) = \alpha_\epsilon(A)\} \\ &= \max\{1/d\left(0, \mathbb{R}_-\cap Y(A-zI)\right) : \\ &\qquad z\in\Lambda_\epsilon(A), \operatorname{Re}(z) = \alpha_\epsilon(A)\} \end{aligned}$$

using our derivative computation before Theorem 3.28. If we take $g(x) = -|x|$ instead, then $\rho_\epsilon = -p$, and

$$\begin{aligned} \text{lip } \rho_\epsilon(A) &\leq \max\{|y| : y \in D^*\Lambda_\epsilon(A \mid z)(-\frac{z}{|z|}), \\ &\qquad z \in \Lambda_\epsilon(A), |z| = \rho_\epsilon(A)\} \\ &= \max\{1/d(0, \mathbb{R}_+(\frac{z}{|z|}) \cap Y(A - zI)) : \\ &\qquad z \in \Lambda_\epsilon(A), |z| = \rho_\epsilon(A)\}. \end{aligned}$$

The upper bounds for lip $\alpha_\epsilon(A)$ and lip $\rho_\epsilon(A)$ obtained above are both not greater than $\text{lip}_\infty\Lambda_\epsilon(A)$ by Proposition 3.29 and so we are done. □

3.7 Resolvent-critical points

Resolvent-critical points are crucial throughout our analysis. They are also, for example, explicitly excluded in the analysis of the quadratic convergence of the algorithm for finding the pseudospectral abscissa in [23]. We investigate their properties further.

Proposition 3.34. *All resolvent-critical points lie in the numerical range of A.*

Proof. Suppose that z is resolvent-critical. Then there exists a right singular vector v of $(A - zI)$ such that $v^H(A - zI)v = 0$, which implies that $v^HAv = zv^Hv = z$ if $|v| = 1$. This means that z lies in the numerical range of A. □

Proposition 3.35. *For ϵ large enough such that $\Lambda_\epsilon(A)$ contains the numerical range of A, $W(A)$, in its interior, the map $\Lambda_\epsilon : M^n \rightrightarrows \mathbb{C}$ is strictly continuous at A for any point in $\Lambda_\epsilon(A)$, and thus Lipschitz continuous at a neighbourhood of A. For α_ϵ and ρ_ϵ to be Lipschitz continuous, we just need the interior of conv$\Lambda_\epsilon(A)$ to contain $W(A)$.*

Proof. For the first part, if $\Lambda_\epsilon(A)$ contains $W(A)$ in its interior, then the points in the boundary of Λ_ϵ are not resolvent-critical by the previous result. Apply Proposition 3.29.

For the second part, by the proof of Corollary 3.33, it suffices to show that if z satisfies $\text{Re } z = \alpha_\epsilon(A)$ and $\underline{\sigma}(A - zI) = \epsilon$, then $z \notin W(A)$. But if z satisfies these conditions, then $z \in \text{conv}\Lambda_\epsilon$. The same goes for ρ_ϵ. □

In Table 1 in page 42, the third example of a 5×5 matrix illustrates that a resolvent-critical can lie outside the convex hull of the spectrum of A. There is a resolvent-critical point close to 0, but the convex hull of the eigenvalues is just $\{-1\}$.

With all that we have done so far, the following is a natural consequence of [12, Corollary 8].

Corollary 3.36. *(to [12, Corollary 8]) Given a matrix A, the set of resolvent-critical values $\{\underline{\sigma}_A(z) \mid z \text{ resolvent critical for } A\}$ is finite.*

Proof. This is just the (semi-algebraic) set of Clarke-critical values of $\underline{\sigma}_A$ by Theorem 3.20, which is finite by [12, Corollary 8]. □

With the above result, we arrive at the following appealing result.

Corollary 3.37. *Given a matrix A, the mappings Λ_ϵ, α_ϵ and ρ_ϵ are Lipschitz around A for all but finitely many $\epsilon \geq 0$, so in particular, for all small $\epsilon > 0$.*

Proof. This is a direct consequence of Theorem 3.26, Corollary 3.36 and Corollary 3.33. □

The conditions that guarantee Lipschitz continuity of the pseudospectral abscissa α_ϵ in the result above are much more general than the conditions in [22, Corollary 8.3]. Firstly, we do not need the assumption that active eigenvalues are nonderogatory made in [22, Corollary 8.3], and our current result holds for all but finitely many ϵ.

Here is another general observation on resolvent-critical points.

Theorem 3.38. *For a fixed A, the set of resolvent-critical points is compact, semi-algebraic with empty interior and contains eigenvalues as isolated points.*

Proof. Denote the set of resolvent-critical points by S_A. The set S_A is bounded by Proposition 3.34. It is clear that S_A is semi-algebraic. As $\underline{\sigma}_A$ is Lipschitz, $\partial^\circ(-\underline{\sigma}_A)$ has closed graph by [31, Proposition 2.1.5(b)] and thus S_A is closed.

Suppose that S_A does not have empty interior. Note that $\underline{\sigma}_A$ has to be constant on a component by Corollary 3.36, and this would mean that $\underline{\sigma}_A$ is constant on a set of nonempty interior, which contradicts the fact that $\underline{\sigma}_A$ cannot have minimizers other than at eigenvalues of A [22, Theorem 4.2]. Thus S_A has empty interior.

Lastly, S_A can be written as a union of curves and points in $\mathbb{C}$. If an eigenvalue, say $\bar{z}$, is not an isolated point in S_A, then it is on some curve. This would mean that $\underline{\sigma}_A$ is zero on a curve, which contradicts the fact that $\underline{\sigma}_A$ is zero only on the set of eigenvalues, which is a finite set. Thus all eigenvalues are isolated in S_A. $\square$

We call $\Lambda'_\epsilon(A) = \{z \mid \underline{\sigma}(A - zI) < \epsilon\}$ the *strict pseudospectrum* of A. The set $\Lambda'_\epsilon(A)$ consists of at most n components (since each component must contain an eigenvalue [86]) and the number of components is clearly a decreasing function of ϵ. There will be some points $\bar{z} \in \mathbb{C}$ where some components meet as ϵ increases. If

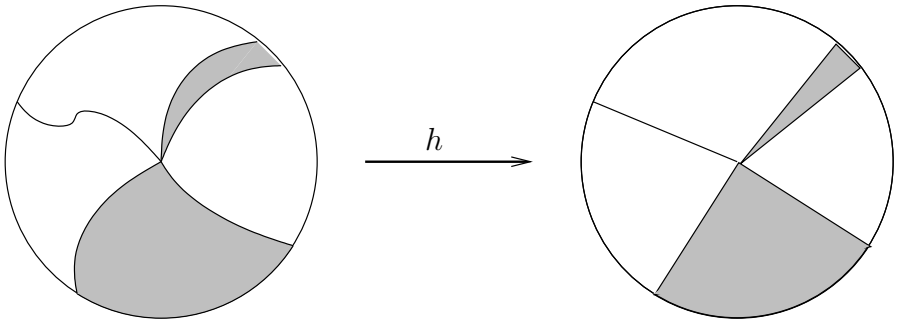
$\Lambda'_\epsilon(A)$ has n components and $\bar{z}$ lies on the boundary of two components of $\Lambda'_\epsilon(A)$, then the distance between A and the set of matrices with repeated eigenvalues is ϵ, and is attained by some matrix $\bar{A}$ having $\bar{z}$ as a repeated eigenvalue. ([1, Theorem 5.1]) It turns out that such points are resolvent-critical as the next theorem will show, generalizing [1, Proposition 4.10].

Theorem 3.39. *If $\bar{z}$ is a common boundary point of components of $\Lambda'_\epsilon(A)$, then $\bar{z}$ is a resolvent-critical point.*

Proof. To reduce notation, let us assume that $\bar{z} = 0$. The rest of the proof will follow by a translation. We look at the structure of $\Lambda_\epsilon(A)$ around 0, where $\epsilon > 0$. Since $\Lambda_\epsilon(A)$ is semi-algebraic, $\Lambda_\epsilon(A)$ is locally conic about 0 by [33, Theorem 4.10]. That is, there is an $r > 0$ and a semi-algebraic homeomorphism

$$h : \Lambda_\epsilon(A) \cap r\mathbb{B} \to [0,1](\Lambda_\epsilon(A) \cap r(\text{bdry } \mathbb{B}))$$

between the two spaces. Since $(\Lambda_\epsilon(A) \cap r(\text{bdry } \mathbb{B}))$ is a finite union of arcs, it follows that the boundary of $\Lambda_\epsilon(A) \cap r\mathbb{B}$ would consist of curves which start from 0 and end at somewhere on $r(\text{bdry } \mathbb{B})$. The diagram below illustrates this.



$\Lambda_\epsilon(A) \cap r\mathbb{B}$ $\qquad$ $[0,1](\Lambda_\epsilon(A) \cap r(\text{bdry}\,\mathbb{B})$

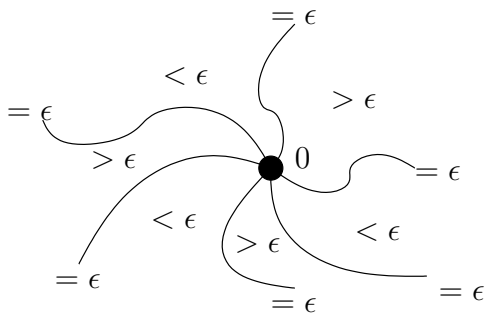
We use a proof by contradiction. Suppose that 0 is not resolvent-critical. Then $0 \notin Y(A)$ and by Proposition 3.22, $\Lambda^c_\epsilon(A)$ is Clarke regular at 0 with normal

cone $N_{\Lambda_\epsilon^c(A)}(0) = \mathbb{R}_+ Y(A)$. Note that $N_{\Lambda_\epsilon^c(A)}(0)$ is pointed, otherwise $0 \in Y(A)$, contradicting the assumption that 0 is not resolvent-critical.

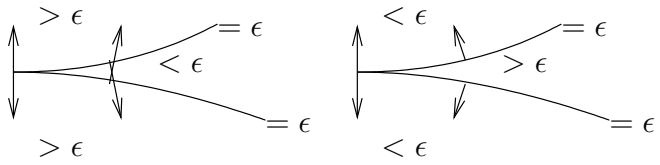
The set $\{z \mid \underline{\sigma}(A - zI) = \epsilon\}$ is semi-algebraic, and has empty interior since the only local minimizers of $\underline{\sigma}_A$ are eigenvalues of A [22, Theorem 4.2], and so it is a union of smooth curves. We now prove that the curves are boundaries of both $\Lambda_\epsilon(A)$ and $\Lambda_\epsilon^c(A)$. By considering the sign of $\underline{\sigma}_A - \epsilon$ on either side of such a curve, we distinguish three cases. In the following diagram, both case 1 and case 2 cannot hold, because local maxima and local minima of $\underline{\sigma}_A$ are resolvent-critical, and this would make 0 resolvent-critical as well, since the set of resolvent-critical points is closed by Theorem 3.38.

$\underline{\sigma}(A - zI) < \epsilon$ $\underline{\sigma}(A - zI) = \epsilon$ $\underline{\sigma}(A - zI) < \epsilon$ 0	$\underline{\sigma}(A - zI) > \epsilon$ $\underline{\sigma}(A - zI) = \epsilon$ $\underline{\sigma}(A - zI) > \epsilon$ 0
Case 1	Case 2

Therefore, the general diagram would be as below, with the value of $\underline{\sigma}_A$ alternating above and below ϵ as we circle the origin, crossing the curves where $\underline{\sigma}_A = \epsilon$.



Two different arcs cannot be tangent at 0 since $N_{\Lambda_\epsilon^c(A)}(0)$ will otherwise not be pointed, as the diagrams below show.



Since $\Lambda_\epsilon^c(A)$ is Clarke regular at 0, its tangent cone $T_{\Lambda_\epsilon^c(A)}(0)$ is convex, so the picture above can contain only one sector where $\underline{\sigma}_A > \epsilon$. It now follows that 0 cannot be the boundary point of two components of $\Lambda'_\epsilon(A)$. This completes the proof. □

If we can prove the following about the pseudospectral abscissa α_ϵ, we would be able to conclude that the pseudospectral abscissa is Lipschitz continuous.

Conjecture 3.40. *The points where the pseudospectral abscissa α_ϵ are attained are not resolvent-critical.*

A natural question to ask after Theorem 3.38 is the following.

Conjecture 3.41. *The number of resolvent-critical points is finite.*

3.8 Acknowledgements

We wish to thank Michael Overton for discussions about much of Section 8, in particular leading to Corollary 8.5. We also thank Diethard Klatte and two anonymous referees for a wide range of comments and suggestions, which greatly improved the article.

CHAPTER 4

LIPSCHITZ BEHAVIOR OF THE ROBUST REGULARIZATION

This section is based on [63], and it contains material in [63] that was not already in [61]. In Section 4.1, we discuss the relationship between calmness and Lipschitz continuity, while in Section 4.2, we discuss the relationship between calmness and robust regularization. In Section 4.3, we prove some results on the robust regularization in the general case. In Section 4.4, we state and prove our main result that at any point $\bar{x}$, the ϵ-robust regularization a semi-algebraic function is Lipschitz at a for all small $\epsilon > 0$. Finally in Section 4.5, we revisit 1-peaceful sets introduced in Section 4.3 and prove some results on how they are relevant to robust regularization.

4.1 Calmness as an extension to Lipschitzness

We begin by discussing the relation between calmness and Lipschitz continuity, which will be important in the proofs in Section 4.4 later. Throughout the chapter, we will limit ourselves to the single-valued case. For more on these topics and their set-valued extensions, we refer the reader to [80].

Definition 4.1. Let $F : X \to \mathbb{R}^m$ be a single-valued map, where $X \subset \mathbb{R}^n$.

(a) [80, Section 8F] Define the *calmness modulus* of F at $\bar{x}$ with respect to X to be

$$\begin{aligned} \operatorname{calm} F(\bar{x}) &:= \inf\{\kappa \mid \text{There is a neighbourhood } V \text{ of } \bar{x} \text{ such that} \\ &\qquad\qquad |F(x) - F(\bar{x})| \le \kappa\,|x - \bar{x}| \text{ for all } x \in V \cap X\} \\ &= \limsup_{x \xrightarrow[X]{} \bar{x}} \frac{|F(x) - F(\bar{x})|}{|x - \bar{x}|}. \end{aligned}$$

Here, $x \underset{X}{\to} \bar{x}$ means that $x \in X$ and $x \to \bar{x}$. The function F is *calm* at $\bar{x}$ with respect to X if $\operatorname{calm} F(\bar{x}) < \infty$.

(b)[80, Definition 9.1] Define the *Lipschitz modulus* of F at $\bar{x}$ with respect to X to be

$$\begin{aligned} \operatorname{lip} F(\bar{x}) &:= \inf\{\kappa \mid \text{There is a neighbourhood } V \text{ of } \bar{x} \text{ such that} \\ &\qquad |F(x) - F(x')| \le \kappa |x - x'| \text{ for all } x, x' \in V \cap X\} \\ &= \limsup_{\substack{x,x' \underset{X}{\to} \bar{x} \\ x \neq x'}} \frac{|F(x) - F(x')|}{|x - x'|}. \end{aligned}$$

The function F is *Lipschitz* at $\bar{x}$ with respect to X if $\operatorname{lip} F(\bar{x}) < \infty$. $\diamond$

As can be seen in the definitions, Lipschitz continuity is a more stringent form of continuity than calmness. In fact, they are related in the following manner.

Proposition 4.2. *Suppose that $F : X \to \mathbb{R}^m$ where $X \subset \mathbb{R}^n$.*

(a) $\limsup_{x \underset{X}{\to} \bar{x}} \operatorname{calm} F(x) \le \operatorname{lip} F(\bar{x})$.

(b) If there is an open set U containing $\bar{x}$ such that $U \cap X$ is convex, then $\operatorname{lip} F(\bar{x}) = \limsup_{x \underset{X}{\to} \bar{x}} \operatorname{calm} F(x)$.

Proof. To simplify notation, let $\kappa := \limsup_{x \underset{X}{\to} \bar{x}} \operatorname{calm} F(x)$.

(a) For any $\epsilon > 0$, we can find a point x_ϵ such that $|\bar{x} - x_\epsilon| < \epsilon$ and $\operatorname{calm} F(x_\epsilon) > \kappa - \epsilon$. Then we can find a point $\tilde{x}_\epsilon$ such that $|x_\epsilon - \tilde{x}_\epsilon| < \epsilon$ and $|F(x_\epsilon) - F(\tilde{x}_\epsilon)| > (\kappa - \epsilon)\,|x_\epsilon - \tilde{x}_\epsilon|$. As ϵ can be made arbitrarily small, we have $\kappa \le \operatorname{lip} F(\bar{x})$ as needed.

(b) For every $\epsilon > 0$, there is some neighborhood of $\bar{x}$, say $\mathbb{B}_\delta(\bar{x})$, such that

$$\operatorname{calm} F(x) \le \kappa + \epsilon \text{ if } x \in \mathbb{B}_\delta(\bar{x}) \cap X.$$

For any $y, z \in \mathbb{B}_\delta(\bar{x}) \cap X$, consider the line segment joining y and z, which we denote by $[y, z]$. As $\operatorname{calm} F(\tilde{x}) \leq \kappa + \epsilon$ for all $\tilde{x} \in [y, z]$, there is a neighborhood around $\tilde{x}$, say $V_{\tilde{x}}$, such that $V_{\tilde{x}} \cap X$ is convex and $|F(\hat{x}) - F(\tilde{x})| \leq (\kappa + 2\epsilon)\,|\hat{x} - \tilde{x}|$ for all $\hat{x} \in V_{\tilde{x}} \cap X$.

As $[y, z]$ is compact, choose finitely many $\tilde{x}$ such that the union of $V_{\tilde{x}}$ covers $[y, z]$. We can add y and z into our choice of points and rename them as $\tilde{x}_1, \ldots, \tilde{x}_k$ in their order on the line segment $[y, z]$, with $\tilde{x}_1 = y$ and $\tilde{x}_k = z$. Also, we can find a point $\hat{x}_i$ between $\tilde{x}_i$ and $\tilde{x}_{i+1}$ such that $\hat{x}_i \in V_{\tilde{x}_i} \cap V_{\tilde{x}_{i+1}}$. Therefore, we add these $\hat{x}_i$ into $\tilde{x}_1, \ldots, \tilde{x}_k$ and get a new set $x_1, \ldots, x_K$, again in their order on the line segment and $x_1 = y$, $x_K = z$.

We have:

$$\begin{aligned} |F(y) - F(z)| &\leq \sum_{i=1}^{K-1} |F(x_i) - F(x_{i+1})| \\ &\leq \sum_{i=1}^{K-1} (\kappa + 2\epsilon)\,|x_i - x_{i+1}| \\ &\leq (\kappa + 2\epsilon)\,|y - z|\,, \end{aligned}$$

and as ϵ is arbitrary, $\operatorname{lip} F(\bar{x}) \leq \kappa$ as claimed. □

Convexity is a strong assumption here, but some analogous condition is needed, as the following examples show.

Example 4.3. (a) Consider the set $X \subset \mathbb{R}$ defined by

$$X = \left(\bigcup_{i=1}^{\infty} \left[\frac{1}{3^i}, \frac{2}{3^i} \right] \right) \cup \{0\}\,,$$

and define the function $F : X \to \mathbb{R}$ by

$$F(x) = \begin{cases} \frac{1}{3^i} & \text{if } \frac{1}{3^i} \leq x \leq \frac{2}{3^i}, \\ 0 & \text{if } x = 0. \end{cases}$$

It is clear that $\operatorname{calm} F(x) = 0$ for all $x \in X \setminus \{0\}$ since F is constant on each component of X, and $\operatorname{calm} F(0) = 1$. But

$$\begin{aligned} \operatorname{lip} F(0) &= \lim_{i\to\infty} \frac{F\left(\frac{1}{3^i}\right) - F\left(\frac{2}{3^{i+1}}\right)}{\frac{1}{3^i} - \frac{2}{3^{i+1}}} \\ &= \lim_{i\to\infty} \frac{\frac{1}{3^i} - \frac{1}{3^{i+1}}}{\frac{1}{3^i} - \frac{2}{3^{i+1}}} \\ &= 2. \end{aligned}$$

Thus, $\limsup_{x\to 0} \operatorname{calm} F(x) < \operatorname{lip} F(0)$.

(b) Consider $X \subset \mathbb{R}^2$ defined by $X := \{(x_1, x_2) \mid x_2^2 = x_1^4\}$ and the function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F(x_1, x_2) = x_2$. One can easily check that $\limsup_{x\to 0} \operatorname{calm} F(x) = 0$ and $\operatorname{lip} F(0,0) = 1$. This is an example of a semi-algebraic function where inequality holds.$\diamond$

Note that $\operatorname{calm} F(\bar{x})$ can be strictly smaller than $\operatorname{lip} F(\bar{x})$ even if X is convex, as demonstrated below.

Example 4.4. (a) Consider $F : \mathbb{R} \to \mathbb{R}$ defined by

$$F(x) = \begin{cases} 0 & \text{if } x = 0, \\ x^2 \sin\left(\frac{1}{x^2}\right) & \text{otherwise.} \end{cases}$$

Here, $\operatorname{calm} F(0) = 0$, but $\operatorname{lip} F(0) = \infty$.

(b) Consider $F : \mathbb{R}^2 \to \mathbb{R}$ defined by:

$$F(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 \le 0 \\ x_1 & \text{if } 0 \le x_1 \le x_2/2 \\ -x_1 & \text{if } 0 \le x_1 \le -x_2/2 \\ 2x_2 & \text{if } x_1 \ge |x_2|/2. \end{cases}$$

We can calculate $\operatorname{calm} F(0,0) = 2/\sqrt{5}$, and $\operatorname{lip} F(0,0) = 2$, so this gives $\operatorname{calm} F(0,0) < \operatorname{lip} F(0,0)$. This is an example of a semi-algebraic function where inequality holds.$\diamond$

At this point, we make a remark about subdifferentially regular functions. It turns out that the calmness and Lipschitz moduli are equal for subdifferentially regular functions.

Proposition 4.5. *If $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is Lipschitz continuous at $\bar{x}$ and subdifferentially regular there, then* $\operatorname{calm} f(\bar{x}) = \operatorname{lip} f(\bar{x})$.

Proof. By [80, Theorem 9.13], $\operatorname{lip} f(\bar{x}) = \max\{|v| \mid v \in \partial f(\bar{x})\}$. If $v \in \partial f(\bar{x})$, then $v \in \hat{\partial} f(\bar{x})$, and we observe that $\operatorname{calm} f(\bar{x}) \geq |v|$ because

$$\begin{aligned} f(\bar{x} + tv) &\geq f(\bar{x}) + \langle v, tv \rangle + o(|t|) \\ &= f(\bar{x}) + |v|\,|tv| + o(|t|). \end{aligned}$$

Therefore $\operatorname{calm} f(\bar{x}) \leq \operatorname{lip} f(\bar{x}) = \max\{|v| \mid v \in \partial f(\bar{x})\} \leq \operatorname{calm} f(\bar{x})$, which implies that all three terms are equal. $\square$

4.2 Calmness and robust regularization

Recall the definition of the robust regularization in Definition 1.1. To study the robust regularization, it is useful to study the dependence of $\bar{f}_\epsilon(x)$ on ϵ instead of on x. For a point $x \in X$, define $g_x : \mathbb{R}_+ \to \mathbb{R}$ by

$$g_x(\epsilon) = \bar{f}_\epsilon(x).$$

To simplify notation, we write $g \equiv g_x$ if it is clear from context. Here are a few basic properties of g_x.

Proposition 4.6. *For $f : X \to \mathbb{R}$ and g_x as defined above, we have the following:*

(a) g_x is monotonically nondecreasing.

(b) If f is continuous in a neighborhood of x, then g_x is continuous in a neighborhood of 0.

Proof. Part (a) is obvious. For part (b), we prove upper and lower semicontinuity separately. We can write $g_x(\epsilon)$ as $\max_{|u|\leq 1} f(x + \epsilon u)$. Since g_x is a maximum of lower semicontinuous functions, g_x is lower semicontinuous.

Next, we prove that g_x is upper semicontinuous. Suppose that $\epsilon_r \to \epsilon$, and $g_x(\epsilon_r) \geq \alpha$. We want to prove that $g_x(\epsilon) \geq \alpha$. By the compactness of the unit ball, there is some u_r such that $g_x(\epsilon_r) = f(x + \epsilon_r u_r)$. For a limit point $\bar{u}$ of $\{u_r\}_r$, we have $g_x(\epsilon) \geq f(x + \epsilon\bar{u}) \geq \alpha$, which is what we need. □

It turns out that calmness of the robust regularization is related to the derivative of g_x.

Proposition 4.7. *If $f : X \to \mathbb{R}$ and $\epsilon > 0$, then $\operatorname{calm} \bar{f}_\epsilon(x) \leq \operatorname{calm} g_x(\epsilon)$. If in addition X contains $\mathbb{B}_{\epsilon'}(x)$ for some $\epsilon' > \epsilon$ and g_x is differentiable at ϵ, then*

$$\operatorname{calm} \bar{f}_\epsilon(x) = \operatorname{calm} g_x(\epsilon) = g_x'(\epsilon).$$

Proof. For the first part, we proceed to show that if $\kappa > \operatorname{calm} g_x(\epsilon)$, then $\kappa \geq \operatorname{calm} \bar{f}_\epsilon(x)$. If $|\tilde{x} - x| < \epsilon$, we have

$$\mathbb{B}_{\epsilon - |\tilde{x}-x|}(x) \subset \mathbb{B}_\epsilon(\tilde{x}) \subset \mathbb{B}_{\epsilon + |\tilde{x}-x|}(x),$$

which implies

$$\bar{f}_{\epsilon - |\tilde{x}-x|}(x) \leq \bar{f}_\epsilon(\tilde{x}) \leq \bar{f}_{\epsilon + |\tilde{x}-x|}(x).$$

Then note that if $\tilde{x}$ is close enough to x, we have

$$\bar{f}_\epsilon(\tilde{x}) \leq \bar{f}_{\epsilon+|\tilde{x}-x|}(x) = g_x(\epsilon + |\tilde{x} - x|) \leq g_x(\epsilon) + \kappa\,|\tilde{x} - x|\,,$$

and similarly

$$\bar{f}_\epsilon(\tilde{x}) \geq \bar{f}_{\epsilon-|\tilde{x}-x|}(x) = g_x(\epsilon - |\tilde{x} - x|) \geq g_x(\epsilon) - \kappa\,|\tilde{x} - x|\,,$$

which tells us that $\left|\bar{f}_\epsilon(\tilde{x}) - \bar{f}_\epsilon(x)\right| \leq \kappa\,|\tilde{x} - x|$, which is what we need.

For the second part, it is clear from the definition of the derivative that $g'_x(\epsilon) = \operatorname{calm} g_x(\epsilon)$. We prove that if $\kappa < g'_x(\epsilon)$, then $\kappa \leq \operatorname{calm} \bar{f}_\epsilon(x)$. By the differentiability of g_x, there is some $\bar{\delta} > 0$ such that for any $0 \leq \delta \leq \bar{\delta}$, we have

$$\begin{aligned} \bar{f}_{\epsilon+\delta}(x) &= g_x(\epsilon + \delta) \\ &> g_x(\epsilon) + \kappa\delta \\ &= \bar{f}_\epsilon(x) + \kappa\delta. \end{aligned}$$

For any $0 \leq \delta \leq \bar{\delta}$, there is some $\tilde{x}_\delta \in \mathbb{B}_{\epsilon+\delta}(x)$ such that $f(\tilde{x}_\delta) = \bar{f}_{\epsilon+\delta}(x)$. Let $\hat{x}_\delta = \frac{\delta}{|\tilde{x}_\delta - x|}(\tilde{x}_\delta - x) + x$. We have $\bar{f}_\epsilon(\hat{x}_\delta) = \bar{f}_{\epsilon+\delta}(x)$, which gives $\bar{f}_\epsilon(\hat{x}_\delta) - \bar{f}_\epsilon(x) > \kappa\delta$. Since $\hat{x}_\delta$ was chosen such that $\delta = |\hat{x}_\delta - x|$, we have $\bar{f}_\epsilon(\hat{x}_\delta) - \bar{f}_\epsilon(x) > \kappa\,|\hat{x}_\delta - x|$, which implies $\kappa \leq \operatorname{calm} \bar{f}_\epsilon(x)$ as needed. □

Remark 4.8. A similar statement can be made for $\epsilon = 0$, except that we change calmness to "calm from above" as defined in [80, Section 8F] in both parts.

We have the following corollary.

Corollary 4.9. *If $f : \mathbb{R}^n \to \mathbb{R}$, $\epsilon > 0$ and g_x is Lipschitz at ϵ, then*

$$\operatorname{calm} \bar{f}_\epsilon(x) \leq \operatorname{lip} g_x(\epsilon) = \sup\{|y| \mid y \in \partial g_x(\epsilon)\}\,.$$

Proof. It is clear that $\operatorname{calm} \bar{f}_\epsilon(x) \leq \operatorname{calm} g_x(\epsilon) \leq \operatorname{lip} g_x(\epsilon)$. The formula $\operatorname{lip} g_x(\epsilon) = \sup\{|y| \mid y \in \partial g_x(\epsilon)\}$ follows from [80, Theorem 9.13, Definition 9.1]. □

In general, the robust regularization is calm.

Proposition 4.10. *For a continuous function $f : X \to \mathbb{R}$, there is an $\bar{\epsilon} > 0$ such that $\bar{f}_\epsilon$ is calm at x for all $0 < \epsilon \leq \bar{\epsilon}$ except on a subset of $(0, \bar{\epsilon}]$ of measure zero.*

Proof. By Proposition 4.6(b), since f is continuous at x, g_x is continuous in $[0, \bar{\epsilon}]$ for some $\bar{\epsilon} > 0$. Since g_x is monotonically nondecreasing, it is differentiable in all $[0, \bar{\epsilon}]$ except for a set of measure zero. The derivative $g'_x(\epsilon)$ equals $\operatorname{calm} \bar{f}_\epsilon(x)$ by Proposition 4.7. □

Remark 4.11. In general, the above result cannot be improved. For an example, let $c : [0, 1] \to [0, 1]$ denote the Cantor function, commonly used in real analysis texts as an example of a function that is not absolutely continuous and not satisfying the Fundamental Theorem of Calculus. Then $\operatorname{calm} \bar{c}_\epsilon(0) = \infty$ for all ϵ lying in the Cantor set.⋄

4.3 Robust regularization in general

In this section, in Corollary 4.15, we prove that if $\operatorname{lip} f(x) < \infty$ for x close to but not equal to $\bar{x}$, then $\operatorname{lip} \bar{f}_\epsilon(\bar{x}) < \infty$ for all small $\epsilon > 0$, even when $\operatorname{lip} f(\bar{x}) = \infty$.

For $F : X \to \mathbb{R}^m$, we may write the robust regularization $F_\epsilon : X \rightrightarrows \mathbb{R}^m$ by $F_\epsilon = F \circ \Phi_\epsilon$, where $\Phi_\epsilon : X \rightrightarrows X$ is defined by $\Phi_\epsilon(x) = \mathbb{B}_\epsilon(x) \cap X$. For reasons that will be clear later in Section 4.5, we consider the extension $\tilde{\Phi}_\epsilon : \mathbb{R}^n \rightrightarrows X$ defined

by $\tilde{\Phi}_\epsilon(x) = \mathbb{B}_\epsilon(x) \cap X$. It is clear that $\tilde{\Phi}_\epsilon \mid_X = \Phi_\epsilon$ using our previous notation, and it follows straight from the definitions that $\operatorname{lip} \Phi_\epsilon(x) \leq \operatorname{lip} \tilde{\Phi}_\epsilon(x)$ for $x \in X$.

Definition 4.12. We say that $X \subset \mathbb{R}^n$ is *peaceful at* $\bar{x} \in X$ if $\operatorname{lip} \Phi_\epsilon(\bar{x})$ is finite for all small $\epsilon > 0$. If in addition $\limsup_{\epsilon \downarrow 0} \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) \leq \kappa$ for all small $\epsilon > 0$, we say that X is peaceful with modulus κ at $\bar{x}$, or *κ-peaceful at* $\bar{x}$.

When $\bar{x}$ lies in the interior of X and ϵ is small enough, then $\tilde{\Phi}_\epsilon$ is Lipschitz with constant 1. In section 4.5, we will find weaker conditions on X for the Lipschitz continuity of $\tilde{\Phi}_\epsilon$. We will see that convex sets are 1-peaceful, but for now, we remark that if X is convex, then Φ_ϵ is globally Lipschitz in X.

Proposition 4.13. *If X is a convex set, then $\Phi_\epsilon(x) \subset \Phi_\epsilon(x') + |x - x'| \, \mathbb{B}$ for all $x, x' \in X$.*

Proof. The condition we are required to prove is equivalent to

$$\mathbb{B}_\epsilon(x) \cap X \subset (\mathbb{B}_\epsilon(x') \cap X) + |x - x'| \, \mathbb{B} \text{ for } x, x' \in X.$$

For any point $\tilde{x} \in \mathbb{B}_\epsilon(x) \cap X$, the line segment $[x', \tilde{x}]$ lies in X, and is of length at most $|\tilde{x} - x| + |x - x'|$. The ball $\mathbb{B}_\epsilon(x')$ can contain the line segment $[x', \tilde{x}]$, in which case $\tilde{x} \in \mathbb{B}_\epsilon(x') \cap X$, or the boundary of $\mathbb{B}_\epsilon(x')$ may intersect $[x', \tilde{x}]$ at a point, say $\hat{x}$. Since X is a convex set, we have $\hat{x} \in \mathbb{B}_\epsilon(x') \cap X$. Furthermore

$$\begin{aligned} |\tilde{x} - \hat{x}| &= |\tilde{x} - x'| - \epsilon \\ &\leq |\tilde{x} - x| + |x - x'| - \epsilon \\ &\leq |x - x'| , \end{aligned}$$

so $\tilde{x} \in (\mathbb{B}_\epsilon(x') \cap X) + |x - x'| \, \mathbb{B}$. □

We remark that if X is nearly radial at $\bar{x}$ as introduced in [61], then X is 1-peaceful: see Section 4.5. The set X is *nearly radial at* $\bar{x}$ if

$$\operatorname{dist}(\bar{x}, x + T_X(x)) \to 0 \text{ as } x \to \bar{x} \text{ in } X.$$

The set X is *nearly radial* if it is nearly radial at all points in X. The notation $T_X(x)$ refers to the *(Bouligand) tangent cone* (or "contingent cone") to X at $x \in X$, formally defined as

$$T_X(\bar{x}) = \{\lim t_r^{-1}(x_r - \bar{x}) : t_r \downarrow 0, \;\; x_r \to \bar{x}, \;\; x_r \in X\}$$

(see, for example, [80, Definition 6.1]). Many sets are nearly radial [61], including for instance semi-algebraic sets, amenable sets and smooth manifolds.

We now present a result on the regularizing property of robust regularization.

Proposition 4.14. *For $F : X \to \mathbb{R}^m$ and $\bar{x} \in X$, suppose that X is peaceful, and there exists a neighborhood U of $\bar{x}$, a convex set $\tilde{X}$, and a function $\tilde{F} : \tilde{X} \to \mathbb{R}^m$ such that $X \cap U \subset \tilde{X} \subset \mathbb{R}^n$, $\tilde{F}|_X = F$ and $\operatorname{lip} \tilde{F}(x) < \infty$. Then $\operatorname{lip} F_\epsilon(\bar{x})$ is finite for all small $\epsilon > 0$.*

Proof. First, we prove that $\operatorname{lip} F : X \to \mathbb{R}_+$ is upper semicontinuous. This result is just a slight modification of the first part of [80, Theorem 9.2], but we include the proof for completeness. Suppose that $x_i \to x$. By the definition of $\operatorname{lip} F$, we can find $x_{i,1}$, $x_{i,2} \in X$ such that

$$\begin{aligned} \frac{|F(x_{i,1}) - F(x_{i,2})|}{|x_{i,1} - x_{i,2}|} &> \operatorname{lip} F(x_i) - |x_i - x|, \\ \text{and } |x_{i,j} - x_i| &< |x_i - x| \text{ for } j = 1, 2. \end{aligned}$$

Taking limits as $i \to \infty$, we see that $x_{i,1}, x_{i,2} \to x$, and it follows that

$$\begin{aligned} \operatorname{lip} F(x) &\geq \limsup_{i \to \infty} \frac{|F(x_{i,1}) - F(x_{i,2})|}{|x_{i,1} - x_{i,2}|} \\ &= \limsup_{i \to \infty} \operatorname{lip} F(x_i). \end{aligned}$$

Thus $\operatorname{lip} F : X \to \mathbb{R}_+$ is upper semicontinuous.

So for ϵ_1 small enough, choose $\epsilon_2 < \epsilon_1$ such that $\operatorname{lip} F$ is bounded above in $C_1 = (\mathbb{B}_{\epsilon_1+\epsilon_2}(\bar{x}) \backslash \mathbb{B}_{\epsilon_1-\epsilon_2}(\bar{x})) \cap X$, say by the constant κ_1. Then for any $\kappa_2 > \kappa_1$ and any $x \in C_1$, there is an ϵ_x such that F is Lipschitz continuous on $\mathbb{B}_{\epsilon_x}(x) \cap X$ with constant κ_2 with respect to X. Thus $\cup_{x \in C_1} \{\mathbb{B}_{\epsilon_x}(x)\}$ is an open cover of C_1.

By the Lebesgue Number Lemma, there is a constant δ such that if x_1, x_2 lie in C_1 and $|x_1 - x_2| \leq \delta$, then the line segment $[x_1, x_2]$ lies in one of the open balls $\mathbb{B}_{\epsilon_x}(x)$ for some $x \in C_1$. We may assume that $\delta < \epsilon_2$.

Also, since X is peaceful at $\bar{x}$, choose ϵ_1 small enough so that $\operatorname{lip} \Phi_{\epsilon_1}(\bar{x})$ is finite, say $\operatorname{lip} \Phi_{\epsilon_1}(\bar{x}) < K$. If X is convex, then this is possible due to Proposition 4.13. We can assume that $K > 2$. Therefore, there is an open set $V \subset U$ about $\bar{x}$ such that Φ_{ϵ_1} is Lipschitz in $V \cap X$ with constant K, that is $\Phi_{\epsilon_1}(x) \subset \Phi_{\epsilon_1}(x') + K\,|x - x'|\,\mathbb{B}$ for all $x, x' \in V \cap X$.

So, for $x, x' \in V \cap \mathbb{B}_{\frac{\delta}{2K}}(\bar{x}) \cap X$, we want to show that

$$F_{\epsilon_1}(x) \subset F_{\epsilon_1}(x') + K\kappa_2\,|x - x'|\,\mathbb{B}.$$

Suppose that $y \in F_{\epsilon_1}(x)$. So $y = F(\tilde{x})$ for some $\tilde{x} \in \mathbb{B}_{\epsilon_1}(x) \cap X$. If $\tilde{x} \in \mathbb{B}_{\epsilon_1 - \frac{\delta}{2K}}(\bar{x})$, then $\tilde{x} \in \mathbb{B}_{\epsilon_1}(x') \cap X$ because $|x' - \bar{x}| \leq \frac{\delta}{2K}$. So $y \in F_{\epsilon_1}(x')$. Otherwise $\tilde{x} \in (\mathbb{B}_{\epsilon_1+\frac{\delta}{2K}}(\bar{x}) \backslash \mathbb{B}_{\epsilon_1-\frac{\delta}{2K}}(\bar{x})) \cap X$.

We have $\Phi_{\epsilon_1}(x) \subset \Phi_{\epsilon_1}(x') + K\,|x - x'|\,\mathbb{B}$. So there is some $\hat{x} \in \Phi_{\epsilon_1}(x')$ such that

$$|\hat{x} - \tilde{x}| \leq K\,|x - x'| \leq K\frac{\delta}{2K} = \frac{\delta}{2}.$$

Furthermore,

$$|\hat{x} - \bar{x}| \leq |\tilde{x} - x| + |x - \bar{x}| + |\hat{x} - \tilde{x}| \leq \epsilon_1 + \frac{\delta}{2K} + \frac{\delta}{2} \leq \epsilon_1 + \frac{3\delta}{4} < \epsilon_1 + \epsilon_2,$$

and

$$|\hat{x} - \bar{x}| \geq |\tilde{x} - x| - |x - \bar{x}| - |\hat{x} - \tilde{x}| \geq \epsilon_1 - \frac{\delta}{2K} - \frac{\delta}{2} \geq \epsilon_1 - \frac{3\delta}{4} > \epsilon_1 - \epsilon_2.$$

Hence $\hat{x} \in (\mathbb{B}_{\epsilon_1+\epsilon_2}(\bar{x}) \backslash \mathbb{B}_{\epsilon_1-\epsilon_2}(\bar{x})) \cap X$. Since $|\hat{x} - \tilde{x}| < \delta$, the line segment $[\hat{x}, \tilde{x}]$ lies in $\mathbb{B}_{\epsilon_x}(x)$ for some $x \in X$. Since the line segment $[\hat{x}, \tilde{x}]$ is convex and $\operatorname{lip} \tilde{F}$ is bounded from above by κ_2 there, we have

$$\begin{aligned} |F(\tilde{x}) - F(\hat{x})| &= \left|\tilde{F}(\tilde{x}) - \tilde{F}(\hat{x})\right| \\ &< \kappa_2 \, |\tilde{x} - \hat{x}| \end{aligned}$$

by [80, Theorem 9.2]. We note that

$$\begin{aligned} F(\tilde{x}) &\in F(\hat{x}) + \kappa_2 \, |\hat{x} - \tilde{x}| \, \mathbb{B} \\ &\subset F_{\epsilon_1}(x') + \kappa_2 \, |\hat{x} - \tilde{x}| \, \mathbb{B} \\ &\subset F_{\epsilon_1}(x') + K\kappa_2 \, |x - x'| \, \mathbb{B}, \end{aligned}$$

and we are done. □

We are now ready to relate $\operatorname{lip} \bar{f}_\epsilon(\bar{x})$ to $\operatorname{lip} f(\bar{x})$. We remind the reader that in the proof of Corollary 4.15 below, $f_\epsilon : X \rightrightarrows \mathbb{R}$ is a set-valued map as introduced in Definition 1.1, which is similar to $\bar{f}_\epsilon$ but maps to intervals in $\mathbb{R}$.

Corollary 4.15. *For $f : X \to \mathbb{R}$, if the conditions in Proposition 4.14 holds (with $F = f$), then $\operatorname{lip} \bar{f}_\epsilon(\bar{x}) < \infty$ for all small $\epsilon > 0$.*

Proof. By Proposition 4.14, we have $\operatorname{lip} f_\epsilon(\bar{x}) < \infty$ with the given conditions. It remains to prove that $\operatorname{lip} \bar{f}_\epsilon(\bar{x}) \leq \operatorname{lip} f_\epsilon(\bar{x})$. We can do this by proving that $\operatorname{lip} \bar{S}(\bar{x}) \leq \operatorname{lip} S(\bar{x})$, where $S : X \rightrightarrows \mathbb{R}$ is a set-valued map, and $\bar{S} : X \to \mathbb{R}$ is defined by $\bar{S}(x) = \sup\{y \mid y \in S(x)\}$. Note that if $S = f_\epsilon$, then $\bar{S} = \overline{(f_\epsilon)} = \bar{f}_\epsilon$.

For any $\kappa > \operatorname{lip} S(x)$, we have $\mathbf{d}(S(\tilde{x}), S(\hat{x})) \leq \kappa |\tilde{x} - \hat{x}|$ for $\tilde{x}, \hat{x} \in X$ close enough to x by [80, Definition 9.26]. The definition of the Pompeiu-Hausdorff distance tells us that $S(\tilde{x}) \subset S(\hat{x}) + \kappa |\tilde{x} - \hat{x}|$, which implies $\bar{S}(\tilde{x}) \leq \bar{S}(\hat{x}) + \kappa |\tilde{x} - \hat{x}|$. By reversing the roles of $\tilde{x}$ and $\hat{x}$, we obtain $\left|\bar{S}(\tilde{x}) - \bar{S}(\hat{x})\right| \leq \kappa |\tilde{x} - \hat{x}|$. So $\kappa > \operatorname{lip} \bar{S}(x)$, and since κ is arbitrary, we have $\operatorname{lip} \bar{S}(x) \leq \operatorname{lip} S(x)$ as needed. □

4.4 Semi-algebraic robust regularization

In this section, in Theorem 4.18, we prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and semi-algebraic, then at any given point, the robust regularization is locally Lipschitz there for all sufficiently small $\epsilon > 0$. This theorem is more appealing than Corollary 4.15 because the required condition is weaker. The condition $\operatorname{lip} f(x) < \infty$ for all x close to but not equal to $\bar{x}$ in Corollary 4.15 is a strong condition because if a function is not Lipschitz at a point $\bar{x}$, it is likely that it is not Lipschitz at some points close to $\bar{x}$ as well. For example in $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = \left|\sqrt{x_1}\right|$, f is not Lipschitz at all points where $x_1 = 0$.

We proceed to prove the main theorem of this section in the steps outlined below.

Proposition 4.16. *For $f : X \to \mathbb{R}$, where $X \subset \mathbb{R}^n$ is convex, define $G : X \times \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ by*

$$G(x, \epsilon) := \limsup_{\tilde{\epsilon} \to \epsilon} \operatorname{lip} \bar{f}_{\tilde{\epsilon}}(x).$$

If f is semi-algebraic, then the maps $(x, \epsilon) \mapsto \operatorname{calm} \bar{f}_\epsilon(x)$, $(x, \epsilon) \mapsto \operatorname{lip} \bar{f}_\epsilon(x)$ and G are semi-algebraic.

Proof. The semi-algebraic nature is a consequence of the Tarski-Seidenberg quan-

tifier elimination. □

The semi-algebraicity of $(x, \epsilon) \mapsto \operatorname{calm} \bar{f}_\epsilon(x)$ gives us an indication of how the map $\epsilon \mapsto \operatorname{calm} \bar{f}_\epsilon(x)$ behaves asymptotically.

Proposition 4.17. *Suppose that $f : X \to \mathbb{R}$ is continuous and semi-algebraic, where $X \subset \mathbb{R}^n$. Fix $x \in X$. Then $\operatorname{calm} \bar{f}_\epsilon(x) = o\left(\frac{1}{\epsilon}\right)$ as $\epsilon \searrow 0$. Hence $\bar{f}_\epsilon$ is calm at x for all small $\epsilon > 0$.*

Proof. The map g_x is semi-algebraic because it can be written as a composition of semi-algebraic maps $\epsilon \mapsto (x, \epsilon) \mapsto \bar{f}_\epsilon(x)$. Thus g_x is differentiable on some open interval of the form $(0, \bar{\epsilon})$ for $\bar{\epsilon} > 0$. Recall that $\operatorname{calm} g_x(\epsilon) = g_x'(\epsilon)$ by Proposition 4.7.

We show that for any $K > 0$, we can reduce $\bar{\epsilon}$ if necessary so that the map $\epsilon \mapsto \operatorname{calm} \bar{f}_\epsilon(x)$ is bounded from above by $\epsilon \mapsto \frac{K}{\epsilon}$ on $\epsilon \in [0, \bar{\epsilon}]$. For any $K > 0$, there exists an $\bar{\epsilon} > 0$ such that either $g_x'(\epsilon) \leq \frac{K}{\epsilon}$ for all $0 < \epsilon < \bar{\epsilon}$, or $g_x'(\epsilon) \geq \frac{K}{\epsilon}$ for all $0 < \epsilon < \bar{\epsilon}$. The latter cannot happen, otherwise for any $0 < \epsilon < \bar{\epsilon}$,

$$\begin{aligned} \bar{f}_\epsilon(x) - f(x) &= \int_0^\epsilon g_x'(s) ds \\ &\geq \int_0^\epsilon \frac{K}{s} ds = \infty. \end{aligned}$$

This contradicts the continuity of g_x. If ϵ is small enough, the derivatives of g_x exist for all small $\epsilon > 0$ and $g_x'(\epsilon) = \operatorname{calm} \bar{f}_\epsilon(x)$ by Proposition 4.7. This gives us the required result. □

Consider $f : [0, 1] \to \mathbb{R}$ defined by $f(x) = x^{1/k}$. Then $g_0(\epsilon) = \epsilon^{1/k}$, so $\operatorname{calm} \bar{f}_\epsilon(0) = g_0'(\epsilon) = \frac{1}{k}\epsilon^{(1/k)-1}$. As $k \to \infty$, we see that the bound above is tight.

We are now ready to state the main theorem of this chapter. In the particular case of $X = \mathbb{R}^n$, we have the following theorem.

Theorem 4.18. *Consider any continuous semi-algebraic function $f : \mathbb{R}^n \to \mathbb{R}$. At any fixed point $\bar{x} \in \mathbb{R}^n$, the robust regularization $\bar{f}_\epsilon$ is Lipschitz at $\bar{x}$, and its calmness and Lipschitz moduli, $\operatorname{calm} \bar{f}_\epsilon(\bar{x})$ and $\operatorname{lip} \bar{f}_\epsilon(\bar{x})$, agree for all sufficiently small ϵ and behave like $o\left(\frac{1}{\epsilon}\right)$ as $\epsilon \downarrow 0$.*

Proof. In view of Proposition 4.17, we only need to prove the there is some $\bar{\epsilon} > 0$ such that $\operatorname{lip} \bar{f}_\epsilon(\bar{x}) = \operatorname{calm} \bar{f}_\epsilon(\bar{x})$ for all $\epsilon \in (0, \bar{\epsilon}]$. We can assume that $g_{\bar{x}}$ is twice continuously differentiable in $(0, \bar{\epsilon}]$. The graph of $G : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}_+$ as defined in Proposition 4.16 is semi-algebraic, so by the decomposition theorem [33, Theorem 6.7], there is a finite partition of semi-algebraic $\mathcal{C}^2$ manifolds $C_1, \ldots, C_l$ such that $G\,|_{C_i}$ is $\mathcal{C}^2$.

If the segment $\{\bar{x}\} \times (0, \bar{\epsilon}]$ lies in a semi-algebraic manifold C_i of full dimension, then

$$\begin{aligned}
\operatorname{lip} \bar{f}_\epsilon(\bar{x}) &= \limsup_{\tilde{x} \to \bar{x}} \operatorname{calm} \bar{f}_\epsilon(\tilde{x}) \text{ (by Proposition 4.2)} \\
&= \limsup_{\tilde{x} \to \bar{x}} g'_{\tilde{x}}(\epsilon) \text{ (by Proposition 4.7)} \\
&= g'_{\bar{x}}(\epsilon) \\
&= \operatorname{calm} \bar{f}_\epsilon(\bar{x}),
\end{aligned}$$

and we have nothing to do. Therefore, assume that the segment is on the boundary of a manifold C_i of full dimension.

Since G is semi-algebraic, the map $\epsilon \mapsto \limsup_{\alpha \to \epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x})$ is semi-algebraic, so we can reduce $\bar{\epsilon} > 0$ as necessary such that either

(1) $\limsup_{\alpha \to \epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x}) < \operatorname{calm} \bar{f}_\epsilon(\bar{x})$ for all $\epsilon \in (0, \bar{\epsilon}]$, or

(2) $\limsup_{\alpha\to\epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x}) = \operatorname{calm} \bar{f}_\epsilon(\bar{x})$ for all $\epsilon \in (0, \bar{\epsilon}]$, or

(3) $\limsup_{\alpha\to\epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x}) > \operatorname{calm} \bar{f}_\epsilon(\bar{x})$ for all $\epsilon \in (0, \bar{\epsilon}]$.

Case (1) cannot hold because $\operatorname{lip} \bar{f}_\epsilon(\bar{x}) \geq \operatorname{calm} \bar{f}_\epsilon(\bar{x})$. Case (2) is what we seek to prove, so we proceed to show that case (3) cannot happen by contradiction.

We can choose $\tilde{\epsilon}, M_1, M_2 > 0$ such that $0 < \tilde{\epsilon} < \bar{\epsilon}$ and

$$\operatorname{calm} \bar{f}_\epsilon(\bar{x}) < M_2 < M_1 < \limsup_{\alpha\to\epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x}) \text{ for all } \epsilon \in [\tilde{\epsilon}, \bar{\epsilon}].$$

We state and prove a lemma important to the rest of the proof before continuing.

Lemma 4.19. *There exists an interval (ϵ_1, ϵ_2) contained in $(\tilde{\epsilon}, \bar{\epsilon}]$ and a manifold $T_1 \subset \mathbb{R}^n \times \mathbb{R}_+$ such that*

(1) $\{\bar{x}\} \times (\epsilon_1, \epsilon_2) \subset \operatorname{cl}(T_1)$.

(2) T_1 is an open $\mathcal{C}^2$ manifold of full dimension.

(3) $H : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}$, defined by $H(x, \epsilon) = \bar{f}_\epsilon(x)$, is $\mathcal{C}^2$ in T_1.

(4) For all $(x, \epsilon) \in T_1$, we have $M_1 \leq g'_x(\epsilon) < \infty$.

(5) $(x, \epsilon) \mapsto g'_x(\epsilon)$ is continuous in T_1.

Proof. Consider the set

$$T := \{(x, \epsilon) \mid M_1 \leq g'_x(\epsilon) < \infty\}.$$

First, we prove that $\{\bar{x}\} \times [\tilde{\epsilon}, \bar{\epsilon}] \subset \operatorname{cl} T$. It suffices to show that for all $\epsilon \in (\tilde{\epsilon}, \bar{\epsilon}]$, $(\bar{x}, \epsilon) \in \operatorname{cl} T$. This can in turn be proven by showing that for all $\delta > 0$, we can find x', ϵ' such that $|\bar{x} - x'| < \delta$, $|\epsilon - \epsilon'| < \delta$ such that $(x', \epsilon') \in T$, or equivalently, $M_1 \leq g'_{x'}(\epsilon') < \infty$.

Since $\limsup_{\alpha\to\epsilon} \operatorname{lip} \bar{f}_\alpha(\bar{x}) > M_1$, there is some ϵ° such that $|\epsilon^\circ - \epsilon| < \frac{\delta}{2}$ and $\operatorname{lip} \bar{f}_{\epsilon^\circ}(\bar{x}) > M_1$.

Next, since

$$\limsup_{x\to\bar{x}} |\partial g_x(\epsilon^\circ)| \geq \limsup_{x\to\bar{x}} \operatorname{calm} \bar{f}_{\epsilon^\circ}(x) = \operatorname{lip} \bar{f}_{\epsilon^\circ}(\bar{x}),$$

there is some x' such that $|\bar{x} - x'| < \delta$ and $|\partial g_{x'}(\epsilon^\circ)| > \frac{1}{2}\operatorname{lip} \bar{f}_{\epsilon^\circ}(\bar{x}) + \frac{1}{2}M_1$.

Finally, since $g_{x'}(\cdot)$ is semi-algebraic, we can find some ϵ' such that $|\epsilon' - \epsilon^\circ| < \frac{\delta}{2}$, $g'_{x'}(\epsilon')$ is well defined and finite, and

$$g'_{x'}(\epsilon') > |\partial g_{x'}(\epsilon^\circ)| - \frac{1}{2}(\operatorname{lip} \bar{f}_{\epsilon^\circ}(\bar{x}) - M_1) > M_1.$$

This choice of x' and ϵ' are easily verified to satisfy the requirements stated.

By the decomposition theorem [33, Theorem 6.7], T can be decomposed into a finite disjoint union of $\mathcal{C}^2$ smooth manifolds $T_1, T_2, \ldots, T_p$ on which H is $\mathcal{C}^2$. Since $\{\bar{x}\} \times [\tilde{\epsilon}, \bar{\epsilon}] \subset \operatorname{cl} T$, there must be some T_i of full dimension and (ϵ_1, ϵ_2) such that $\{\bar{x}\} \times (\epsilon_1, \epsilon_2) \subset \operatorname{cl} T_i$. Without loss of generality, let one such T_i be T_1.

Conditions (1), (2), (3) and (4) are automatically satisfied. Note that $g'_x(\epsilon)$ is exactly the derivative of $H(\cdot, \cdot)$ with respect to the second coordinate, and so Property (5) is satisfied. This concludes the proof of the lemma. □

We now continue with the rest of the proof of the theorem. Note that the manifold T_1 is of dimension at least two.

Using Lemma 4.22 which we will prove later, we can construct the map $\varphi : [0, 1) \times (\hat{\epsilon}_1, \hat{\epsilon}_2) \to \operatorname{cl} T_1$, such that its derivative with respect to the second variable exists and is continuous, and $\varphi(0, \epsilon) = (\bar{x}, \epsilon)$ for all $\epsilon \in (\hat{\epsilon}_1, \hat{\epsilon}_2)$.

For each $0 < \delta < 1$, consider the path $\tilde{x}_\delta : [\hat{\epsilon}_1, \hat{\epsilon}_2] \to \mathbb{R}^n$ defined by $\tilde{x}_\delta(\epsilon) := \varphi(\delta, \epsilon)$. We have

$$\begin{aligned} & \bar{f}_{\hat{\epsilon}_2}(\tilde{x}_\delta(\hat{\epsilon}_2)) - \bar{f}_{\hat{\epsilon}_1}(\tilde{x}_\delta(\hat{\epsilon}_1)) \\ = & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \nabla H(\tilde{x}_\delta(s), s) \cdot (\tilde{x}'_\delta(s), 1) ds \\ = & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \nabla_x H(\tilde{x}_\delta(s), s) \cdot \tilde{x}'_\delta(s) ds + \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \nabla_s H(\tilde{x}_\delta(s), s) ds, \end{aligned}$$

where $H(x, \epsilon) = \bar{f}_\epsilon(x)$. The second component of $\nabla H(\tilde{x}_\delta(s), s)$ is simply $g'_{\tilde{x}_\delta(s)}(s)$. The first component can be analyzed as follows:

$$\begin{aligned} & \nabla_x H(\tilde{x}_\delta(s), s) \cdot \tilde{x}'_\delta(s) \\ = & \lim_{t \to 0} \frac{1}{t} (H(\tilde{x}_\delta(s) + t\tilde{x}'_\delta(s), s) - H(\tilde{x}_\delta(s), s) \\ = & \lim_{t \to 0} \frac{1}{t} (\bar{f}_s(\tilde{x}_\delta(s) + t\tilde{x}'_\delta(s)) - \bar{f}_s(\tilde{x}_\delta(s))). \end{aligned}$$

Provided that $t\,|\tilde{x}'_\delta(s)| < s$, $\mathbb{B}_{s - t|\tilde{x}'_\delta(s)|}(\tilde{x}_\delta(s)) \subset \mathbb{B}_s(\tilde{x}_\delta(s) + t\tilde{x}'_\delta(s))$, and so

$$\begin{aligned} & \nabla_x H(\tilde{x}_\delta(s), s) \cdot \tilde{x}'_\delta(s) \\ \geq & \lim_{t \to 0} \frac{1}{t} (\bar{f}_{s - t|\tilde{x}'_\delta(s)|}(\tilde{x}_\delta(s)) - \bar{f}_s(\tilde{x}_\delta(s))) \\ = & |\tilde{x}'_\delta(s)| \lim_{t \to 0} \frac{1}{t\,|\tilde{x}'_\delta(s)|} (\bar{f}_{s - t|\tilde{x}'_\delta(s)|}(\tilde{x}_\delta(s)) - \bar{f}_s(\tilde{x}_\delta(s))) \\ = & -|\tilde{x}'_\delta(s)|\, g'_{\tilde{x}_\delta(s)}(s). \end{aligned}$$

Hence,

$$\begin{aligned} & \bar{f}_{\hat{\epsilon}_2}(\tilde{x}_\delta(\hat{\epsilon}_2)) - \bar{f}_{\hat{\epsilon}_1}(\tilde{x}_\delta(\hat{\epsilon}_1)) \\ = & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \nabla_x H(\tilde{x}_\delta(s), s) \cdot \tilde{x}'_\delta(s) ds + \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \nabla_s H(\tilde{x}_\delta(s), s) ds \\ \geq & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} (1 - |\tilde{x}'_\delta(s)|) g'_{\tilde{x}_\delta(s)}(s) ds. \end{aligned}$$

Since the derivatives of φ are continuous, $\tilde{x}'_\delta(s) \to \tilde{x}'_0(s) = 0$ as $\delta \to 0$ for $\hat{\epsilon}_1 < s < \hat{\epsilon}_2$. In fact, the term $|\tilde{x}'_\delta(s)|$ converges to zero uniformly in $[\hat{\epsilon}_1, \hat{\epsilon}_2]$. To see this,

recall that $\tilde{x}'_\delta(s)$ is a partial derivative of φ. Since φ is $\mathcal{C}^1$, $\tilde{x}'_\delta(s)$ is continuous with respect to s and δ. For any $\beta > 0$ and $s \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$, there exists γ_s such that

$$|\tilde{x}'_\delta(\tilde{s})| < \beta \text{ if } \delta < \gamma_s \text{ and } |\tilde{s} - s| < \gamma_s.$$

The existence of γ such that

$$|\tilde{x}'_\delta(s)| < \beta \text{ if } \delta < \gamma \text{ and } s \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$$

follows by the compactness of $[\hat{\epsilon}_1, \hat{\epsilon}_2]$. So we may choose δ small enough so that

$$(1 - |\tilde{x}'_\delta(s)|) > \frac{M_1 + M_2}{2M_1} \text{ for all } s \in [\hat{\epsilon}_1, \hat{\epsilon}_2].$$

Now, for δ small enough and $i = 1, 2$, we have $g'_{\bar{x}}(\hat{\epsilon}_i) < M_2$, so this gives us $\operatorname{calm} \bar{f}_{\hat{\epsilon}_i}(\bar{x}) = g'_{\bar{x}}(\hat{\epsilon}_i) < M_2$ by Proposition 4.7. Therefore, if δ is small enough,

$$\left|\bar{f}_{\hat{\epsilon}_i}(\tilde{x}_\delta(\hat{\epsilon}_i)) - \bar{f}_{\hat{\epsilon}_i}(\bar{x})\right| \leq M_2 \left|\tilde{x}_\delta(\hat{\epsilon}_i) - \bar{x}\right|.$$

Recall that if the derivative $g'_{\bar{x}}(\epsilon)$ exists, then $g'_{\bar{x}}(\epsilon) = \operatorname{calm} \bar{f}_\epsilon(\bar{x})$ by Proposition 4.7. On the one hand, we have

$$\bar{f}_{\hat{\epsilon}_2}(\bar{x}) - \bar{f}_{\hat{\epsilon}_1}(\bar{x}) = \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} g'_{\bar{x}}(s)ds \leq \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} M_2 ds = M_2(\hat{\epsilon}_2 - \hat{\epsilon}_1).$$

But on the other hand, $\tilde{x}_\delta(s) \in T_1$ for $0 < \delta < 1$, and so $g'_{\tilde{x}_\delta(s)}(s) \geq M_1$ by Lemma

4.19. If δ is small enough, we have

$$\begin{aligned}
& \left|\bar{f}_{\hat{\epsilon}_2}(\bar{x}) - \bar{f}_{\hat{\epsilon}_1}(\bar{x})\right| \\
\geq\ & \left|\bar{f}_{\hat{\epsilon}_2}(\tilde{x}_\delta(\hat{\epsilon}_2)) - \bar{f}_{\hat{\epsilon}_1}(\tilde{x}_\delta(\hat{\epsilon}_1))\right| \\
& \quad - \left(\left|\bar{f}_{\hat{\epsilon}_2}(\tilde{x}_\delta(\hat{\epsilon}_2)) - \bar{f}_{\hat{\epsilon}_2}(\bar{x})\right| + \left|\bar{f}_{\hat{\epsilon}_1}(\tilde{x}_\delta(\hat{\epsilon}_1)) - \bar{f}_{\hat{\epsilon}_1}(\bar{x})\right|\right) \\
\geq\ & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} (1 - |\tilde{x}'_\delta(s)|) g'_{\tilde{x}_\delta(s)}(s) ds \\
& \quad - M_2 \left(|\tilde{x}_\delta(\hat{\epsilon}_2) - \bar{x}| + |\tilde{x}_\delta(\hat{\epsilon}_1) - \bar{x}|\right) \\
\geq\ & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} (1 - |\tilde{x}'_\delta(s)|) M_1 ds - M_2(|\tilde{x}_\delta(\hat{\epsilon}_2) - \bar{x}| + |\tilde{x}_\delta(\hat{\epsilon}_1) - \bar{x}|) \\
\geq\ & \int_{\hat{\epsilon}_1}^{\hat{\epsilon}_2} \left(\frac{M_1 + M_2}{2}\right) ds - M_2(|\tilde{x}_\delta(\hat{\epsilon}_2) - \bar{x}| + |\tilde{x}_\delta(\hat{\epsilon}_1) - \bar{x}|) \\
=\ & \left(\frac{M_1 + M_2}{2}\right) (\hat{\epsilon}_2 - \hat{\epsilon}_1) - M_2(|\tilde{x}_\delta(\hat{\epsilon}_2) - \bar{x}| + |\tilde{x}_\delta(\hat{\epsilon}_1) - \bar{x}|).
\end{aligned}$$

As δ is arbitrarily small and the terms $|\tilde{x}_\delta(\hat{\epsilon}_i) - \bar{x}| \to 0$ as $\delta \to 0$ for $i = 1, 2$, we have $\left|\bar{f}_{\hat{\epsilon}_2}(\bar{x}) - \bar{f}_{\hat{\epsilon}_1}(\bar{x})\right| \geq \left(\frac{M_1+M_2}{2}\right)(\hat{\epsilon}_2 - \hat{\epsilon}_1)$. This is a contradiction, and thus we are done. □

Before we prove Lemma 4.22 below, we need to recall the definition of simplicial complexes from [34, Section 3.2.1]. A *simplex* with vertices $a_0, \ldots, a_d$ is

$$\begin{aligned}
[a_0, \ldots, a_d] \ =\ & \{x \in \mathbb{R}^n \mid \exists \lambda_0, \ldots, \lambda_d \in [0, 1], \\
& \qquad \sum_{i=0}^{d} \lambda_i = 1 \text{ and } x = \sum_{i=0}^{d} \lambda_i a_i.\}
\end{aligned}$$

The corresponding *open simplex* is

$$\begin{aligned}
(a_0, \ldots, a_d) \ =\ & \{x \in \mathbb{R}^n \mid \exists \lambda_0, \ldots, \lambda_d \in (0, 1), \\
& \qquad \sum_{i=0}^{d} \lambda_i = 1 \text{ and } x = \sum_{i=0}^{d} \lambda_i a_i.\}
\end{aligned}$$

We shall denote by $\mathrm{int}(\sigma)$ the open simplex corresponding to the simplex σ. A face of the simplex $\sigma = [a_0, \ldots, a_d]$ is a simplex $\tau = [b_0, \ldots, b_e]$ such that

$$\{b_0, \ldots, b_e\} \subset \{a_0, \ldots, a_d\}.$$

A *finite simplicial complex* in $\mathbb{R}^n$ is a finite collection $K = \{\sigma_1, \ldots, \sigma_p\}$ of simplices $\sigma_i \subset \mathbb{R}^n$ such that, for every $\sigma_i, \sigma_j \in K$, the intersection $\sigma_i \cap \sigma_j$ is either empty or is a common face of σ_i and σ_j. We set $|K| = \cup_{\sigma_i \in K} \sigma_i$; this is a semi-algebraic subset of $\mathbb{R}^n$. We recall a result on relating semi-algebraic sets to simplicial complexes.

Theorem 4.20. *[34, Theorem 3.12] Let $S \subset \mathbb{R}^n$ be a compact semi-algebraic set, and $S_1, \ldots, S_p$, semi-algebraic subsets of S. Then there exists a finite simplicial complex K in $\mathbb{R}^n$ and a semi-algebraic homeomorphism $h : |K| \to S$, such that each S_k is the image by h of a union of open simplices of K.*

We need yet another result for the proof of Lemma 4.22.

Proposition 4.21. *Suppose that $\phi : (0,1)^2 \to \mathbb{R}$, not necessarily semi-algebraic, is continuous in $(0,1)^2$. Let $\operatorname{gph} \phi \subset (0,1)^2 \times \mathbb{R}$ be the graph of ϕ. Then for any $t \in (0,1)$, $\operatorname{cl}(\operatorname{gph} \phi) \cap (0,t) \times \mathbb{R}$ is either a single point or a connected line segment.*

Proof. Suppose that $((0,t), a_1)$ and $((0,t), a_2)$ lie in $\operatorname{cl}(\operatorname{gph} \phi)$. We need to show that for any $\alpha \in (a_1, a_2)$, $((0,t), \alpha)$ lies in $\operatorname{cl}(\operatorname{gph} \phi)$.

For any $\epsilon > 0$, we can find points $p_1, p_2 \in (0,1)^2$ such that the points $(p_1, \tilde{a}_1), (p_2, \tilde{a}_2) \in \operatorname{gph} \phi$ are such that $|\tilde{a}_i - a_i| < \epsilon$ and $|p_i - (0,t)| < \epsilon$ for $i = 1, 2$. Recall that by definition $\tilde{a}_i = \phi(p_i)$ for $i = 1, 2$. Choose ϵ such that $\tilde{a}_1 + \epsilon < \tilde{a}_2 - \epsilon$. By the intermediate value theorem, for any $\alpha \in (\tilde{a}_1 + \epsilon, \tilde{a}_2 - \epsilon)$, there exists a point p in the line segment $[p_1, p_2]$ such that $\phi(p) = \alpha$. Moreover, $|p - (0,t)| < \max_{i=1,2} |p_i - (0,t)|$. Letting $\epsilon \to 0$, we see that $((0,t), \alpha) \in \operatorname{cl}(\operatorname{gph} \phi)$ as needed. □

We now prove our last result important for the proof of Theorem 4.18. The

proof of the lemma below is similar to the proof of the Curve Selection Lemma in [34, Theorem 3.13].

Lemma 4.22. *Let $S \subset \mathbb{R}^n$ be a semi-algebraic set, and $\tau : [\epsilon_1, \epsilon_2\} \to \mathbb{R}^n$ be a semi-algebraic curve such that $\tau((\epsilon_1, \epsilon_2]) \cap S = \emptyset$ and $\tau([\epsilon_1, \epsilon_2]) \subset \mathrm{cl}\,(S)$. Then there exists a function $\varphi : [0,1] \times [\hat{\epsilon}_1, \hat{\epsilon}_2] \to \mathbb{R}^n$, with $[\hat{\epsilon}_1, \hat{\epsilon}_2] \neq \emptyset$ and $[\hat{\epsilon}_1, \hat{\epsilon}_2] \subset [\epsilon_1, \epsilon_2]$, such that*

(1) $\varphi(0, \epsilon) = \tau(\epsilon)$ for $\epsilon \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$ and $\varphi((0,1] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]) \subset S$.

(2) The partial derivative of φ with respect to the second variable, which we denote by $\frac{\partial}{\partial \epsilon}\varphi$, exists and is continuous in $[0,1] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$.

Proof. Replacing S with its intersection with a closed bounded set containing $\tau([\epsilon_1, \epsilon_2])$, we can assume S is bounded. Then $\mathrm{cl}(S)$ is a compact semi-algebraic set. By Theorem 4.20, there is a finite simplicial complex K and a semi-algebraic homeomorphism $h : |K| \to \mathrm{cl}(S)$, such that S and $\tau([\epsilon_1, \epsilon_2])$ are images by h of a union of open simplices in K . In particular, this means that there is an open interval $(\hat{\epsilon}_1, \hat{\epsilon}_2) \subset [\epsilon_1, \epsilon_2]$ such that $\tau((\hat{\epsilon}_1, \hat{\epsilon}_2))$ is an image by h of a 1-dimensional open simplex in K. Since $h^{-1} \circ \tau((\hat{\epsilon}_1, \hat{\epsilon}_2))$ is in $\mathrm{cl}(S)$ but not in S, there is a simplex σ of K which has $h^{-1} \circ \tau([\hat{\epsilon}_1, \hat{\epsilon}_2])$ lying in the boundary of σ, and $h(\mathrm{int}(\sigma)) \subset S$.

Let $\hat{\sigma}$ be the barycenter of σ. Define the map $\delta : [0,1] \times [\hat{\epsilon}_1, \hat{\epsilon}_2] \to \mathbb{R}^n$ by

$$\delta(t, \epsilon) = (1-t)h^{-1} \circ \tau(\epsilon) + t\hat{\sigma}.$$

The map above satisfies $\delta((0,1] \times (\hat{\epsilon}_1, \hat{\epsilon}_2)) \subset \mathrm{int}(\sigma)$. By contracting the interval $[\hat{\epsilon}_1, \hat{\epsilon}_2]$ slightly, $\varphi = h \circ \delta$ satisfies property (1).

By contracting the interval $[\hat{\epsilon}_1, \hat{\epsilon}_2]$ if necessary and applying the decomposition theorem [33, Theorem 6.7], we can assume that φ is $\mathcal{C}^1$ in the set $(0, \bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$ for

some $\bar{t} \in (0,1)$.

Since τ is semi-algebraic, we contract the interval $[\hat{\epsilon}_1, \hat{\epsilon}_2]$ again if necessary so that τ is $\mathcal{C}^1$ there. Therefore, $\frac{\partial}{\partial \epsilon}\varphi$ exists in $[0,\bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$. It remains to show that $\frac{\partial}{\partial \epsilon}\varphi$ is continuous in $[0,\bar{t}]\times[\hat{\epsilon}_1, \hat{\epsilon}_2]$. We do this by showing that $\frac{\partial}{\partial \epsilon}\varphi_i : [0,\bar{t}]\times[\hat{\epsilon}_1, \hat{\epsilon}_2] \to \mathbb{R}$, the ith component of the derivative with respect to the second variable, is continuous for each i.

Since $\frac{\partial}{\partial \epsilon}\varphi_i$ is continuous in $(0,\bar{t}]\times[\hat{\epsilon}_1, \hat{\epsilon}_2]$, it remains to show that it is continuous at every point in $\{0\} \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$. The graph of $\frac{\partial}{\partial \epsilon}\varphi_i$ corresponding to the domain $(0,\bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$, which we denote by $\text{gph}\left(\frac{\partial}{\partial \epsilon}\varphi_i\right)$, is a subset of $(0,\bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2] \times \mathbb{R}$. We show that $((0,\epsilon), \frac{\partial}{\partial \epsilon}\varphi_i(0,\epsilon)) \in \text{cl}\left(\text{gph}\left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right)$. For small $t_1, t_2 > 0$, consider $\varphi_i(t_1, \epsilon - t_2)$ and $\varphi_i(t_1, \epsilon + t_2)$. By the intermediate value theorem, there is some $\tilde{\epsilon} \in (\epsilon - t_2, \epsilon + t_2)$ such that

$$\frac{\partial}{\partial \epsilon}\varphi_i(t_1, \tilde{\epsilon}) = \frac{1}{2t_2}(\varphi_i(t_1, \epsilon + t_2) - \varphi_i(t_1, \epsilon - t_2.))$$

If t_2 were chosen such that

$$\left|\frac{1}{2t_2}(\varphi_i(0, \epsilon + t_2) - \varphi_i(0, \epsilon - t_2)) - \frac{\partial}{\partial \epsilon}\varphi_i(0, \epsilon)\right|$$

is small and t_1 is chosen such that

$$\left|\frac{1}{2t_2}(\varphi_i(t_1, \epsilon + t_2) - \varphi_i(t_1, \epsilon - t_2)) - \frac{1}{2t_2}(\varphi_i(0, \epsilon + t_2) - \varphi_i(0, \epsilon - t_2))\right|$$

is small, then $\left|\frac{\partial}{\partial \epsilon}\varphi_i(t_1, \tilde{\epsilon}) - \frac{\partial}{\partial \epsilon}\varphi_i(0, \epsilon)\right|$ is small. Taking $t_2 \to 0$ and $t_1 \to 0$, we have $((0,\epsilon), \frac{\partial}{\partial \epsilon}\varphi_i(0,\epsilon)) \in \text{cl}\left(\text{gph}\left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right)$ as desired.

Recall that the graph $\text{gph}\left(\frac{\partial}{\partial \epsilon}\varphi_i\right)$ is taken corresponding to the domain $(0,\bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$, and is a manifold of dimension 2 in $\mathbb{R}^3$. Its boundary is of dimension 1 [34, Proposition 3.16], so the intersection of $\text{cl}\left(\text{gph}\left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right)$ with $\{0\} \times [\hat{\epsilon}_1, \hat{\epsilon}_2] \times \mathbb{R}$ is of dimension 1 as well, and is homeomorphic to a closed line segment. There

cannot be an interval $[\tilde{\epsilon}_1, \tilde{\epsilon}_2] \subset [\hat{\epsilon}_1, \hat{\epsilon}_2]$ on which cl $\left(\text{gph } \left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right) \cap \{0\} \times \{\epsilon\} \times \mathbb{R}$ has more than one value for all $\epsilon \in [\tilde{\epsilon}_1, \tilde{\epsilon}_2]$ because by appealing to Proposition 4.21, this implies that the dimension cannot be 1. We note however that it is possible that there exists an $\bar{\epsilon} \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$ such that cl $\left(\text{gph } \left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right) \cap \{0\} \times \{\bar{\epsilon}\} \times \mathbb{R}$ is a 1-dimensional line segment. This can only happen for only finitely many $\bar{\epsilon} \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$ due to semi-algebraicity.

In any case, we can contract the interval $[\hat{\epsilon}_1, \hat{\epsilon}_2]$ if necessary so that cl $\left(\text{gph } \left(\frac{\partial}{\partial \epsilon}\varphi_i\right)\right) \cap \{0\} \times \{\epsilon\} \times \mathbb{R}$ is a single point for all $\epsilon \in [\hat{\epsilon}_1, \hat{\epsilon}_2]$. This means that for any $(t, \tilde{\epsilon}) \to (0, \epsilon)$, we have $\frac{\partial}{\partial \epsilon}\varphi_i(t, \tilde{\epsilon}) \to \frac{\partial}{\partial \epsilon}\varphi_i(0, \epsilon)$, establishing the continuity of $\frac{\partial}{\partial \epsilon}\varphi_i(\cdot, \cdot)$ on $[0, \bar{t}] \times [\hat{\epsilon}_1, \hat{\epsilon}_2]$. A reparametrization allows us to assume that $\bar{t} = 1$, and we are done. □

4.5 1-peaceful sets

In this section, we prove that $X \subset \mathbb{R}^n$ is nearly radial implies X is 1-peaceful using the Mordukhovich Criterion [80, Theorem 9.40], which relates the Lipschitz modulus of set-valued maps to normal cones of its graph. The next section discusses further properties of nearly radial sets and how they are common in analysis.

The Mordukhovich Criterion requires the domain of the set-valued map to be $\mathbb{R}^n$, so we recall the map $\tilde{\Phi}_\epsilon : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ by $\tilde{\Phi}_\epsilon(x) = \mathbb{B}_\epsilon(x) \cap X$. Recall that $\tilde{\Phi}_\epsilon|_X = \Phi_\epsilon$ and $\operatorname{lip} \Phi_\epsilon(x) \leq \operatorname{lip} \tilde{\Phi}_\epsilon(x)$ for all $x \in X$.

We now present our result on the relation between 1-peaceful sets and nearly radial sets.

Theorem 4.23. *If X is nearly radial at $\bar{x}$ and locally closed there, then X is*

1-*peaceful at* $\bar{x}$. *The converse holds if* X *is Clarke regular for all points in a neighborhood around* $\bar{x}$.

Proof. The graph of $\tilde{\Phi}_\epsilon$ is the intersection of $\mathbb{R}^n \times X$ and the set $D \subset \mathbb{R}^n \times \mathbb{R}^n$ defined by

$$D := \{(x, y) \mid \|x - y\| \leq \epsilon\}.$$

By applying a rule on the normal cones of products of sets [80, Proposition 6.41], we infer that $N_{\mathbb{R}^n \times X}(x, y) = \{\mathbf{0}\} \times N_X(y)$. Define the real valued function $g_0 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ by $g_0(x, y) := \frac{1}{2}\|x - y\|^2$. Then the gradient of g_0 is $\nabla g_0(x, y) = (x - y, y - x)$.

From this point, we assume that $\|x - y\| = \epsilon$. The normal cone of D at (x, y) is $N_D(x, y) = \mathbb{R}_+\{(x - y, y - x)\}$ using [80, Exercise 6.7]. On applying a rule on the normal cones of intersections [80, Theorem 6.42], we get

$$N_{\operatorname{gph} \tilde{\Phi}_\epsilon}(x, y) \subset (\{\mathbf{0}\} \times N_X(y)) + \mathbb{R}_+\{(x - y, y - x)\}. \tag{4.5.1}$$

Furthermore, if X is Clarke regular at y, the above set inclusion is an equation. Since X is locally closed at $\bar{x}$, $\tilde{\Phi}_\epsilon$ is locally closed at $\bar{x}$ if ϵ is small enough. By the Mordukhovich criterion , $\tilde{\Phi}_\epsilon$ has the Aubin Property at (x, y) if and only if the graphical modulus $\operatorname{lip} \tilde{\Phi}_\epsilon(x \mid y)$ is finite. It can be calculated by appealing to the

formulas for the coderivative D^* and outer norm $|\cdot|^+$ below.

$$\begin{aligned}
\operatorname{lip}\tilde{\Phi}_\epsilon(x \mid y) &= \left|D^*\tilde{\Phi}_\epsilon(x \mid y)\right|^+ \text{ (by [80, Theorem 9.40])} \\
&= \sup_{w\in\mathbb{B}} \sup_{z\in D^*\tilde{\Phi}_\epsilon(w)} \|z\| \text{ (by [80, Section 9D])} \\
&= \sup\left\{\|z\| \mid (w,z) \in \operatorname{gph} D^*\tilde{\Phi}_\epsilon, \|w\| \le 1\right\} \\
&= \sup\left\{\|z\| \mid (-z,w) \in N_{\operatorname{gph}\tilde{\Phi}_\epsilon}(x,y), \|w\| \le 1\right\} \\
&\quad \text{(by [80, Definition 8.33])} \\
&\le \sup\Big\{ \|z\| \mid (-z,w) \in (\{\mathbf{0}\} \times N_X(y)) \\
&\qquad +\mathbb{R}_+\{(x-y, y-x)\}, \|w\| \le 1.\Big\}
\end{aligned} \tag{4.5.2}$$

We can assume that $z = y - x$ with a rescaling, and $w = y - x + v$ for some $v \in N_X(y)$. Since $(\{\mathbf{0}\} \times N_X(y)) + \mathbb{R}_+\{(x-y, y-x)\}$ is positively homogeneous set, we could find the supremum of $\frac{\|z\|}{\|w\|}$ in the same set and the formula reduces to

$$\begin{aligned}
\operatorname{lip}\tilde{\Phi}_\epsilon(x \mid y) &\le \sup_{v\in N_X(y)} \frac{\|y-x\|}{\|y-x+v\|} \\
&= \sup_{v\in N_X(y)} \frac{\|x-y\|}{\|(x-y)-v\|} \\
&= \frac{\|x-y\|}{d(x-y, N_X(y))}.
\end{aligned} \tag{4.5.3}$$

For a fixed $x \neq y$, say $\bar{x}$, we have $1/\operatorname{lip}\tilde{\Phi}_\epsilon(\bar{x} \mid y) \ge \frac{d(\bar{x}-y, N_X(y))}{\|\bar{x}-y\|}$. First, we prove that for any open set W about $\bar{x}$, we have

$$\inf_{\substack{y\in W\cap X \\ y\neq\bar{x}}} \frac{d(\bar{x}-y, N_X(y))}{\|\bar{x}-y\|} = \inf_{\substack{y\in W\cap X \\ y\neq\bar{x}}} \frac{d(\bar{x}-y, \hat{N}_X(y))}{\|\bar{x}-y\|}. \tag{4.5.4}$$

It is clear that "$\le$" holds because $\hat{N}_X(y) \subset N_X(y)$, so we proceed to prove the other inequality. Consider $d(\bar{x}-y, N_X(y))$. Let $v \in P_{N_X(y)}(\bar{x}-y)$, the projection of $(\bar{x}-y)$ onto $N_X(y)$. Then $v \in N_X(y)$, and so there exists $y_i \to y$, with $y_i \in W \cap X$,

and $v_i \to v$ such that $v_i \in \hat{N}_X(y_i)$. So

$$\begin{aligned} d(\bar{x}-y, N_X(y)) &= d(\bar{x}-y, \mathbb{R}_+(v)) \\ &= \lim_{i\to\infty} d(\bar{x}-y, \mathbb{R}_+(v_i)) \\ &= \lim_{i\to\infty} d(\bar{x}-y_i, \mathbb{R}_+(v_i)) \\ &\geq \limsup_{i\to\infty} d(\bar{x}-y_i, \hat{N}_X(y_i)) \\ \Rightarrow \frac{d(\bar{x}-y, N_X(y))}{\|\bar{x}-y\|} &\geq \limsup_{i\to\infty} \frac{d(\bar{x}-y_i, \hat{N}_X(y_i))}{\|\bar{x}-y_i\|}. \end{aligned}$$

Thus equation 4.5.4 holds. Therefore

$$\liminf_{y\to\bar{x}} \frac{d(\bar{x}-y, \hat{N}_X(y))}{\|\bar{x}-y\|} \geq 1 \text{ implies } \limsup_{y\to\bar{x}} \operatorname{lip} \tilde{\Phi}_{\|\bar{x}-y\|}(\bar{x} \mid y) \leq 1,$$

so we may now consider only regular normal cones.

By the Moreau decomposition of the polar cones $\hat{N}_X(y)$ and $\hat{N}_X(y)^*$, we have

$$d(\bar{x}-y, \hat{N}_X(y))^2 + d(\bar{x}-y, \hat{N}_X(y)^*)^2 = \|\bar{x}-y\|^2 \text{ for } y \in X.$$

Since $T_X(y)^* = \hat{N}_X(y)$ always [80, Theorem 6.28(a)], we have

$$d(\bar{x}-y, \hat{N}_X(y))^2 + d(\bar{x}-y, T_X(y)^{**})^2 = \|\bar{x}-y\|^2 \text{ for } y \in X.$$

As $T_X(y) \subset T_X(y)^{**}$ [80, Corollary 6.21], this implies that

$$d(\bar{x}-y, \hat{N}_X(y))^2 + d(\bar{x}-y, T_X(y))^2 \geq \|\bar{x}-y\|^2 \text{ for } y \in X. \tag{4.5.5}$$

Note that if X is nearly radial at $\bar{x}$, then $\frac{1}{\|\bar{x}-y\|} d(\bar{x}-y, T_X(y)) \to 0$ as $\epsilon = \|\bar{x}-y\| \downarrow 0$, $y \in X$. This means that

$$1/\operatorname{lip} \tilde{\Phi}_{\|\bar{x}-y\|}(\bar{x} \mid y) \geq \frac{1}{\|\bar{x}-y\|} d(\bar{x}-y, \hat{N}_X(y)) \to 1,$$

so

$$\limsup_{y \xrightarrow[X]{} \bar{x}, y\neq\bar{x}} \operatorname{lip} \tilde{\Phi}_{\|\bar{x}-y\|}(\bar{x} \mid y) \leq 1,$$

where $y \underset{X}{\to} \bar{x}$ means $y \in X$ and $y \to \bar{x}$.

Recall that $\tilde{\Phi}_\epsilon$ has closed graph, and hence it is outer semicontinuous [80, Theorem 5.7(a)]. It is also locally bounded, so

$$\operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) = \max_{y \in S_\epsilon(\bar{x})} \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x} \mid y)$$

by [72, Theorem 1.42]. This gives us $\limsup_{\epsilon \to 0} \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) \leq 1$, or X is 1-peaceful at $\bar{x}$, as needed.

If we assume that X is Clarke regular in a neighborhood of $\bar{x}$, then Formula (4.5.5) is an equation. Furthermore, (4.5.1), (4.5.2) and (4.5.3) are all equations. Thus if $\lim_{\epsilon \to 0} \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) = 1$, then

$$\frac{1}{\|\bar{x} - y\|} d(\bar{x} - y, \hat{N}_X(y)) = 1/\operatorname{lip} \tilde{\Phi}_{\|\bar{x}-y\|}(\bar{x} \mid y) \to 1 \text{ as } y \underset{X}{\to} \bar{x},\ y \neq \bar{x}.$$

and we have $\frac{1}{\|\bar{x}-y\|} d(\bar{x} - y, T_X(y)) \to 0$ as $y \underset{X}{\to} \bar{x}$ and $y \neq \bar{x}$, which means that X is nearly radial at $\bar{x}$. □

Finally, 1−peaceful sets are interesting in robust regularization for another reason. The Lipschitz modulus of the robust regularization over 1-peaceful sets have Lipschitz modulus bounded above by that of the original function, as the following result shows.

Proposition 4.24. *If X is 1-peaceful and $F : X \to \mathbb{R}^n$ is locally Lipschitz at $\bar{x}$, then*

$$\limsup_{\epsilon \to 0} \operatorname{lip} F_\epsilon(\bar{x}) \leq \operatorname{lip} F(\bar{x}).$$

Proof. We use a set-valued chain rule [80, Exercise 10.39]. Recall the formula $F_\epsilon = (F \circ \tilde{\Phi}_\epsilon) \mid_X$. The mapping $(x, u) \mapsto \tilde{\Phi}_\epsilon(x) \cap F^{-1}(u)$ is locally bounded

because the map $x \mapsto \tilde{\Phi}_\epsilon(x)$ is locally bounded. Thus

$$\operatorname{lip} F_\epsilon(\bar{x}) \leq \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) \cdot \max_{x \in \tilde{\Phi}_\epsilon(\bar{x})} \operatorname{lip} F(x).$$

By Theorem 4.23, $\lim_{\epsilon \to 0} \operatorname{lip} \tilde{\Phi}_\epsilon(\bar{x}) \leq 1$. Also, since $\operatorname{lip} F : \mathbb{R}^n \to \mathbb{R}_+$ is upper semicontinuous, $\limsup_{\epsilon \to 0} \max_{x \in \tilde{\Phi}_\epsilon(\bar{x})} \operatorname{lip} F(x) \leq \operatorname{lip} F(\bar{x})$. Taking limits to both sides gives us what we need. □

Acknowledgement. We thank Mike Todd for his comments that lead to the current statement of Theorem 4.18, and to the two anonymous referees for comments that improved the article.

CHAPTER 5

CONTINUITY OF SET-VALUED MAPS REVISITED IN THE LIGHT OF TAME GEOMETRY

In this chapter, we revisit the following result recorded in [80, Theorem 5.55] and [7, Theorem 1.4.13], and attributed to [60, 30, 83]. The domain of the set-valued map S below can be taken to be a complete metric space, while the range can be taken to be a complete separable metric space, but we shall only state the result in the finite dimensional case. Recall that a set is *nowhere dense* if its closure has empty interior, and *meager* if it is the union of countably many sets that are nowhere dense in X.

Theorem 5.1. *Let $X \subset \mathbb{R}^n$ and $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a closed-valued set-valued map. Assume S is either outer semicontinuous or inner semicontinuous relative to X. Then the set of points $x \in X$ where S fails to be continuous relative to X is meager in X.*

The map

$$S(x) := \begin{cases} 0 & \text{if } x \text{ is rational} \\ 1 & \text{if } x \text{ is irrational} \end{cases} \tag{5.0.1}$$

shows that it is possible for a set-valued map to be nowhere outer and nowhere inner semicontinuous.

The following example shows the sharpness of Theorem 5.1, if we move to noncomplete spaces.

Example 5.2. Let $c_{00}(\mathbb{N})$ denote the vector space of all real sequences $x = \{x_n\}_{n\in\mathbb{N}}$ with finite support $\operatorname{supp}(x) := \{i \in \mathbb{N} : x_i \neq 0\}$. Then the operator

$S_1(x) = \text{supp}(x)$ is everywhere inner semicontinuous and nowhere outer semicontinuous, while the operator $S_2(x) = \mathbb{Z}\backslash S_1(x)$ is everywhere outer semicontinuous and nowhere inner semicontinuous.

A stronger concept of continuity for set-valued maps is that of *strict continuity* [80, Definition 9.28], which is equivalent to Lipschitz continuity when the map is single-valued. For set-valued maps $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with bounded values, strict continuity is quantified by the Hausdorff distance. Namely, a set-valued map S is strictly continuous at $\bar{x}$ (relative to X) if the quantity

$$\text{lip}_X S(\bar{x}) := \limsup_{\substack{x,\, x' \to \bar{x} \\ x \neq x'}} \frac{\mathbf{d}(S(x), S(x'))}{|x - x'|}$$

is bounded. In the general case (that is, when S maps to unbounded sets), we say that S is strictly continuous, whenever the truncated map $S_r : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined by

$$S_r(x) := S(x) \cap r\,\mathbb{B},$$

is Lipschitz continuous for every $r > 0$.

Another natural question to ask is whether we can extend continuity in Theorem 5.1 to strict continuity. The following result shows that a strictly increasing continuous single-valued map from the reals to the reals could be non-Lipschitz everywhere, let alone satisfy the Aubin property.

Example 5.3. Let $A \subset \mathbb{R}$ be a measurable set with the property that for every $a, b \in \mathbb{R}$, $a < b$, the Lebesgue measure of $A \cap (a, b)$ satisfies $0 < m(A \cap [a, b]) < |b - a|$. Consider the function $f : [0, 1] \to \mathbb{R}$ defined by $f(x) = m(A \cap (0, x))$. Note that the derivative $f'(x)$ exists almost everywhere and is equal to $\chi_A(x)$, the characteristic function of A (equal to 1 if $x \in A$ and 0 if not). This means that

every point $\bar{x} \in [0,1]$ is arbitrarily close to a point x where $f'(x)$ is well-defined and equals zero. Thus $(0,1) \in N_{\mathrm{gph}\,(f)}(\bar{x}, f(\bar{x}))$. The function f is strictly increasing, so it has a continuous inverse $g : [0, f(1)] \to [0,1]$. Applying the Mordukhovich criterion, we obtain that g does not have the Aubin property at $f(\bar{x})$. It follows that g is not strictly continuous at $f(\bar{x})$ and in fact neither is so at any $y \in [0, f(1)]$.

We mention an important property of semialgebraic (more generally, o-minimal) sets. A set is (topologically) *generic* if its complement is meagre. Genericity and full measure (*i.e.,* almost everywhere) are different ways to affirm that a given property holds in a large set. However, these notions are often complementary. In particular, it is possible for a generic subset of $\mathbb{R}^n$ to be of null measure, or for a full measure set to be meager (see [76] for example). Nonetheless, this situation disappears in our setting.

Proposition 5.4. *[Genericity in a semialgebraic setting] Let U, V be semialgebraic subsets of $\mathbb{R}^n$, and assume $V \subset U$. Then the following properties are equivalent:*

(i) V is dense in U ;

(ii) V is (topologically) generic in U ;

(iii) $U \setminus V$ is of null (Lebesgue) measure ;

(iv) the dimension of $U \setminus V$ is strictly smaller than that of U.

In Section 5.1, we prove a Sard-type result for local Pareto minima. In Sections 5.2, 5.3 and 5.4, we prove some preparatory results for our main theorem in Section 5.5 on the generic strict continuity of semi-algebraic set-valued maps. We conclude with some applications and observations of our result in Section 5.6.

5.1 A Sard result for local (Pareto) minima

In this section we use simple properties on the continuity of set-valued maps to obtain a Sard type result for local minima for both scalar and vector-valued functions. Let us recall that a (single-valued) function $f : X \to \mathbb{R}$ is called *lower semicontinuous at* $\bar{x}$ if

$$\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x}) .$$

The function f is called *lower semicontinuous*, if it is lower semicontinuous at every $x \in X$. It is well-known that a function f is lower semicontinuous if and only if its sublevel sets

$$\mathrm{lev}_{\leq r} f := \{x \in X : f(x) \leq r\}$$

are closed for all $r \in \mathbb{R}$.

Proposition 5.5. *[Sublevel map] Let D be a closed subset of a complete metric space X and $f : D \to \mathbb{R}$ be a lower semicontinuous function. Then the (sublevel) set-valued map*

$$\begin{cases} L_f : \mathbb{R} \rightrightarrows D \\ L_f(r) = \mathrm{lev}_{\leq r} f \cup \partial D \end{cases}$$

is outer semicontinuous. Moreover, L_f is continuous at $\bar{r} \in f(D)$ if and only if there is no $x \in \mathrm{int}\,(D)$ such that $f(x) = \bar{r}$ and x is a local minimizer of f.

Proof. The map $L'_f : \mathbb{R} \rightrightarrows D$ defined by $L'_f(r) = f^{-1}((-\infty, r])$ is outer semicontinuous since f is lower semicontinuous (see [80, Example 5.5] for example), so L_f is easily seen to be outer semicontinuous.

We now prove that L_f is inner semicontinuous at $\bar{r}$ under the additional conditions mentioned in the statement. For any $r_i \to \bar{r}$, we want to show that if $\bar{x} \in L_f(\bar{r})$, then there exists $x_i \to \bar{x}$ such that $x_i \in L_f(r_i)$. We can assume that

$f(\bar{x}) = \bar{r}$ and $r_i < \bar{r}$ for all i, otherwise we can take $x_i = \bar{x}$ for i large enough. Since $\bar{x}$ is not a local minimum, for any $\epsilon > 0$, there exists $\delta > 0$ such that if $|\bar{r} - r_i| < \delta$, there exists an x_i such that $f(x_i) \leq r_i$ and $|x_i - \bar{x}| < \epsilon$.

For the converse, assume now that L_f is inner semicontinuous at $\bar{r}$. Then taking $r_i \nearrow \bar{r}$ we obtain that for every $x \in \text{int}(D) \cap f^{-1}(\bar{r})$, there exists $x_i \in f^{-1}(r_i)$ with $x_i \to x$. Since $f(x_i) = r_i < \bar{r} = f(x)$, x cannot be a local minimum. □

According to the above result, if f has no local minima, then the set-valued map L_f is continuous everywhere. The above result has the following interesting consequence.

Corollary 5.6. *[Local minimum values] Let M_f denote the set of local minima of a lower semicontinuous function $f : D \to \mathbb{R}$ (where D is a closed subset of a complete space X). Then the set $f(M_f \cap \text{int}\,(D))$ is meager in $\mathbb{R}$.*

Proof. Since the set-valued map L_f (defined in Proposition 5.5) is outer semicontinuous (with closed-values), it is generically continuous by Theorem 5.1. The second part of Proposition 5.5 yields the result on f. □

It is interesting to compare the above result with the classical Sard theorem. We recall that the Sard theorem asserts that the image of critical points (derivative not surjective) of a C^k function $f : \mathbb{R}^n \to \mathbb{R}^m$ is of measure zero provided $k > n-m$. (See [81]; the case $m = 1$ is known as the Sard-Brown theorem [20].) Corollary 5.6 asserts the topological sparsity of the (smaller) set of minimum values for scalar functions ($m = 1$), without assuming anything but lower semicontinuity (and completeness of the domain).

We shall now extend Corollary 5.6 in the vectorial case. We recall that a set $K \subset \mathbb{R}^m$ is a *cone*, if $\lambda K \subset K$ for all $\lambda \geq 0$. A cone K is called *pointed* if $K \cap (-K) = \{\mathbf{0}_m\}$ (or equivalently, if K contains no full lines). It is well-known that there is a one-to-one correspondence between pointed convex cones of $\mathbb{R}^m$ and partial orderings in $\mathbb{R}^m$. In particular, given such a cone K of $\mathbb{R}^m$ we set $y_1 \leq_K y_2$ if and only if $y_2 - y_1 \in K$ (see for example, [80, Section 3E]). Further, given a set-valued map $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ we say that

- $\bar{x}$ is a *(local) Pareto minimum* of S with *(local) Pareto minimum value* $\bar{y}$ if there is a neighborhood U of $\bar{x}$ such that if $x \in U$ and $y \in S(x)$, then $y \not\leq_K \bar{y}$, i.e., $S(U) \cap (\bar{y} - K) = \bar{y}$.

For $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, define the map $S_K : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ by $S_K(x) = S(x) + K$. The graph of S_K is also known as the *epigraph* [45, 56] of S. One easily checks that $y \in S_K(x)$ implies $y + K \subset S_K(x)$. Here is our result on local Pareto minimum values.

Proposition 5.7. *[Pareto minimum values] Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be an outer semicontinuous map such that $y \in S(x)$ implies $y + K \subset S(x)$ (that is, $S = S_K$). Then the set of local Pareto minimum values is meager.*

Proof. Since S is outer semicontinuous, then S^{-1} is outer semicontinuous as well by [80, Theorem 5.7(a)], so S^{-1} is generically continuous by Theorem 5.1. Suppose that $\bar{y}$ is a local Pareto minimum of a local Pareto minimizer $\bar{x}$.

By the definition of local Pareto minimum, there is a neighborhood U of $\bar{x}$ such that if $y \leq_K \bar{y}$ and $y \neq \bar{y}$, then $S^{-1}(y) \cap U = \emptyset$. (We can assume that y is arbitrarily close to $\bar{y}$ since $S^{-1}(y) \subset S^{-1}(\lambda y + (1 - \lambda)\bar{y})$ for all $0 \leq \lambda \leq 1$.)

Therefore, $\bar{x} \notin \liminf_{y \to \bar{y}} S^{-1}(y)$. In other words, S^{-1} is not continuous at $\bar{y}$. Therefore, the set of local Pareto minimum values is meager. $\square$

We show how the above result compares to critical point results. Let us recall from [52] the definition of critical points of a set-valued map. Given a metric space X (equipped with a distance ρ) we denote by $B_\rho(x, \lambda)$ the set of all $x' \in X$ such that $\rho(x, x') \leq \lambda$.

Definition 5.8. Let (X, ρ_1) and (Y, ρ_2) be metric spaces, and let $S : X \rightrightarrows Y$. For $(x, y) \in \operatorname{gph}(S)$, we set

$$\operatorname{Sur} S(x \mid y)(\lambda) = \sup \{r \geq 0 \mid B_{\rho_2}(y, r) \subset S(B_{\rho_1}(x, \lambda))\}$$

and then for $(\bar{x}, \bar{y}) \in \operatorname{gph}(S)$ define the *rate of surjection* of S at $(\bar{x}, \bar{y})$ by

$$\operatorname{sur} S(\bar{x} \mid \bar{y}) = \liminf_{(x,y,\lambda) \to (\bar{x}, \bar{y}, +0)} \frac{1}{\lambda} \operatorname{Sur} S(x \mid y)(\lambda).$$

We say that S is *critical* at $(\bar{x}, \bar{y}) \in \operatorname{gph}(S)$ if $\operatorname{sur} S(\bar{x} \mid \bar{y}) = 0$, and regular otherwise. Also, $\bar{y}$ is a *(proper) critical value* of S if there exists $\bar{x}$ such that $\bar{y} \in S(\bar{x})$ and S is critical at $(\bar{x}, \bar{y})$.

This definition of critical values characterizes the values at which metric regularity is absent. In the particular case where $S : \mathbb{R}^n \to \mathbb{R}^m$ is a $\mathcal{C}^1$ function, critical points correspond exactly to where the Jacobian has rank less than m. We refer to [52] for more details.

One easily sees that if y is a Pareto minimum value of S, then there exists $x \in X$ such that $(x, y) \in \operatorname{gph}(S)$, and $\operatorname{Sur} S(x \mid y)(\lambda) = 0$ for all small $\lambda > 0$. This readily implies that y is a critical value.

5.2 Extending the Mordukhovich criterion and a critical value result

The two results of this subsection are important ingredients of the forthcoming proof of our main theorem. The first result we need is an adaptation of the Mordukhovich criterion (Proposition 2.11) to the case where the domain of a set-valued function S is (included in) a smooth submanifold $\mathcal{X}$ of $\mathbb{R}^n$. (Note that this new statement recovers the Mordukhovich criterion if $\mathcal{X} = \mathbb{R}^n$.)

Proposition 5.9. *(Extended Mordukhovich criterion) Let $\mathcal{X} \subset \mathbb{R}^n$ be a $\mathcal{C}^1$ smooth submanifold of dimension d and $S : \mathcal{X} \rightrightarrows \mathbb{R}^m$ be a set-valued map whose graph is locally closed at $(\bar{x}, \bar{y}) \in \operatorname{gph}(S)$. Consider the mapping*

$$\begin{cases} H : \mathbb{R}^m \rightrightarrows \mathbb{R}^n \\ H(y^*) = D^*S(\bar{x} \mid \bar{y})(y^*) \cap T_{\mathcal{X}}(\bar{x}). \end{cases}$$

If $H(\mathbf{0}_m) = \{\mathbf{0}_n\}$, or equivalently

$$N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}}(\bar{x}) \times \{\mathbf{0}_m\}) = \{\mathbf{0}_{n+m}\},$$

then S has the Aubin property at $\bar{x}$ for $\bar{y}$ relative to $\mathcal{X}$. Furthermore,

$$\operatorname{lip}_{\mathcal{X}} S(\bar{x} \mid \bar{y}) = |H|^+ = \sup \left\{ \frac{|u|}{|v|} \mid (u, v) \in N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}}(\bar{x}) \times \mathbb{R}^m) \right\}.$$

Proof. Fix $(\bar{x}, \bar{y}) \in \operatorname{gph}(S)$ and denote by $N_{\mathcal{X}}(\bar{x})$ the normal space of $\mathcal{X}$ at $\bar{x}$ (seing as subspace of $\mathbb{R}^n$, that is, $T_{\mathcal{X}}(\bar{x}) \oplus N_{\mathcal{X}}(\bar{x}) = \mathbb{R}^n$). Given a closed neighborhood U of $(\bar{x}, \bar{y})$, we define the function

$$\begin{cases} \tilde{S} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m \\ \operatorname{gph}(\tilde{S}) = (\operatorname{gph}(S) \cap U) + (N_{\mathcal{X}}(\bar{x}) \times \{\mathbf{0}_m\}). \end{cases}$$

Shrinking the neighborhood U around $(\bar{x}, \bar{y})$ if necessary, we may assume that every $(x, y) \in U$ can be represented uniquely as a sum of elements in $(\mathcal{X} \times \mathbb{R}^m) \cap U$ and $N_{\mathcal{X}}(\bar{x}) \times \{\mathbf{0}_m\}$. Since $\operatorname{gph}(S)$ is locally closed, we can choose U small enough so that $\operatorname{gph}(S) \cap U$ is closed. Further, since $\operatorname{gph}(\tilde{S})$ is homeomorphic to $(\operatorname{gph}(S) \cap U) \times \mathbb{R}^{n-d}$, it is also closed.

Step 1: (Relating $\tilde{S}$ to H) By applying a result on the normal cones under set addition [80, Exercise 6.44], we have $N_{\operatorname{gph}(\tilde{S})}(\bar{x}, \bar{y}) \subset N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}}(\bar{x}) \times \mathbb{R}^m)$. To prove the reverse inclusion, note that for every $(x, y) \in \operatorname{gph}(\tilde{S})$ near $(\bar{x}, \bar{y})$ with $(x, y) = (x_1, y) + (x_2, \mathbf{0}_m)$, where $(x_1, y) \in \operatorname{gph}(S)$ and $x_2 \in N_{\mathcal{X}}(\bar{x})$, one easily sees that $\hat{N}_{\operatorname{gph}(\tilde{S})}(x, y) \supset \hat{N}_{\operatorname{gph}(S)}(x_1, y) \cap (T_{\mathcal{X}}(\bar{x}) \times \mathbb{R}^m)$. The extension of this inclusion to limiting normal cones is immediate. Therefore we obtain

$$N_{\operatorname{gph}(\tilde{S})}(\bar{x}, \bar{y}) = N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}}(\bar{x}) \times \mathbb{R}^m),$$

and so $D^*\tilde{S}(\bar{x} \mid \bar{y})$ equals the set-valued map H described in the statement. Thus

$$\begin{aligned} D^*\tilde{S}(\bar{x} \mid \bar{y})(\mathbf{0}_m) &= \left\{x^* \mid (x^*, \mathbf{0}_m) \in N_{\operatorname{gph}(\tilde{S})}(\bar{x}, \bar{y})\right\} \\ &= \left\{x^* \mid (x^*, \mathbf{0}_m) \in N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}}(\bar{x}) \times \mathbb{R}^m)\right\} \\ &= \{\mathbf{0}_n\}, \end{aligned}$$

and by the Mordukhovich criterion, the map $\tilde{S}$ has the Aubin property at $\bar{x}$ for $\bar{y}$.

Taking neighborhoods V of $\bar{x}$ and W of $\bar{y}$ so that $S(x) \cap W = \tilde{S}(x) \cap W$ for all $x \in V \cap \mathcal{X}$, we deduce that S has the Aubin property at $\bar{x}$ for $\bar{y}$ relative to $\mathcal{X}$ as asserted.

Step 2: ($\operatorname{lip}_{\mathcal{X}} S(\bar{x} \mid \bar{y}) = |H|^+$) The Mordukhovich criterion on $\tilde{S}$ yields

$$|H|^+ = \operatorname{lip} \tilde{S}(\bar{x} \mid \bar{y}) \geq \operatorname{lip}_{\mathcal{X}} S(\bar{x} \mid \bar{y}).$$

Our task is thus to prove that the above inequality is actually an equality. Since $\operatorname{lip} \tilde{S}(\bar{x} \mid \bar{y}) = |H|^+$, for any $\kappa < |H|^+$ and neighborhoods V of $\bar{x}$ and W of $\bar{y}$, there exist $x_1, x_2 \in V$ such that

$$\tilde{S}(x_2) \cap W \not\subset \tilde{S}(x_1) + \kappa |x_1 - x_2| \mathbb{B}.$$

Note that $\tilde{S}(x_1) = \tilde{S}(P(x_1))$, $\tilde{S}(x_2) = \tilde{S}(P(x_2))$ and $|P(x_1) - P(x_2)| \leq |x_1 - x_2|$, where P stands for the projection of $\mathbb{R}^n$ onto $\bar{x} + T_{\mathcal{X}}(\bar{x})$. We may choose V to be a ball containing $\bar{x}$, and define the projection parametrization $L : (\bar{x} + T_{\mathcal{X}}(\bar{x})) \cap V \to \mathcal{X}$ of the manifold $\mathcal{X}$ by the relation $x - L(x) \in N_{\mathcal{X}}(\bar{x})$. Shrinking V if needed, the map L becomes single-valued and smooth (in fact, it is a local chart of $\mathcal{X}$ at $\bar{x}$ provided we identify $\bar{x} + T_{\mathcal{X}}(\bar{x})$ with $\mathbb{R}^d$). Furthermore, L has Lipschitz constant 1 at $\bar{x}$. Therefore, for any $\epsilon > 0$, we can reduce V as needed so that L is Lipschitz continuous in its domain with Lipschitz constant at most $(1 + \epsilon)$ using standard arguments (*e.g.* [80, Thms 9.7, 9.2]). This means that

$$S(L(x_2)) \cap W = \tilde{S}(x_2) \cap W \not\subset \tilde{S}(x_1) + \kappa |x_1 - x_2| \mathbb{B} = S(L(x_1)) + \kappa |x_1 - x_2| \mathbb{B}.$$

By the Lipschitz continuity of L, we have $|L(x_1) - L(x_2)| \leq (1 + \epsilon) |x_1 - x_2|$, which gives

$$S(L(x_2)) \cap W \not\subset S(L(x_2)) + \frac{\kappa}{(1 + \epsilon)} |L(x_1) - L(x_2)| \mathbb{B},$$

yielding

$$\frac{\kappa}{1 + \epsilon} \leq \operatorname{lip}_{\mathcal{X}} S(\bar{x} \mid \bar{y}).$$

Since κ and ϵ are arbitrary, we conclude that $|H|^+ = \operatorname{lip}_{\mathcal{X}} S(\bar{x} \mid \bar{y})$ as asserted.

The proof is complete. □

The second result is an adaptation of part of [52, Theorem 6]. Recall that for a smooth function $f : \mathbb{R}^n \to \mathbb{R}^m$, $\bar{x} \in \mathbb{R}^n$ is a *critical point* if the derivative $\nabla f(\bar{x})$

is not surjective, while $\bar{y} \in \mathbb{R}^m$ is a *critical value* if there is a critical point $\bar{x}$ for which $f(\bar{x}) = \bar{y}$. (Note this is a particular case of the general definition given in Definition 5.8.)

Lemma 5.10. *Let $\mathcal{X}$ be a $\mathcal{C}^k$ smooth manifold in $\mathbb{R}^n$ of dimension d, and $\mathcal{M}$ be a $\mathcal{C}^k$ manifold in $\mathbb{R}^{n+m}$ such that $\mathcal{M} \subset \mathcal{X} \times \mathbb{R}^m$, with $k > \dim \mathcal{M} - \dim \mathcal{X}$. Then the set of points $x \in \mathcal{X}$ such that there exists some y satisfying $(x, y) \in \mathcal{M}$ and $N_{\mathcal{M}}(x, y) \cap (T_{\mathcal{X}}(x) \times \{\mathbf{0}_m\}) \supsetneq \{\mathbf{0}_{n+m}\}$ is of Lebesgue measure zero in $\mathcal{X}$.*

Proof. Let $\mathrm{Proj}_{\mathcal{M}}$ denote the restriction to the manifold $\mathcal{M}$ of the projection of $\mathcal{X} \times \mathbb{R}^m$ onto $\mathcal{X}$. As $k > \dim \mathcal{M} - \dim \mathcal{X}$, the set of critical values of $\mathrm{Proj}_{\mathcal{M}}$ is a set of measure zero by the classical Sard theorem [81]. Let $(x, y) \in \mathcal{M}$ and assume $(x^*, \mathbf{0}_m) \in N_{\mathcal{M}}(x, y) \cap (T_{\mathcal{X}}(x) \times \{\mathbf{0}_m\})$ with $x^* \neq \mathbf{0}_n$. This gives

$$T_{\mathcal{M}}(x, y) = (N_{\mathcal{M}}(x, y))^{\perp} \subset \{x^*\}^{\perp} \times \mathbb{R}^m,$$

where $\{x^*\}^{\perp} = \{x' \in \mathbb{R}^n \mid \langle x^*, x' \rangle = 0\}$. Since $T_{\mathcal{M}}(x, y) \subset T_{\mathcal{X}}(x) \times \mathbb{R}^m$ we obtain

$$T_{\mathcal{M}}(x, y) \subset (\{x^*\}^{\perp} \cap T_{\mathcal{X}}(x)) \times \mathbb{R}^m.$$

Let Z stand for the subspace on the right hand side. Then the projection of Z onto $T_{\mathcal{X}}(x)$ is a proper subspace of $T_{\mathcal{X}}(x)$. All the more, this applies to $T_{\mathcal{M}}(x, y)$. By [52, Corollary 3], this implies that (x, y) is a singular point of $\mathrm{Proj}_{\mathcal{M}}$, so x is a critical value of $\mathrm{Proj}_{\mathcal{M}}$. The conclusion of the lemma follows. □

5.3 Some preliminary results

We finish this section with two useful results. The first one is well-known (with elementary proof) and is mentioned for completeness.

Proposition 5.11. *If $\mathcal{K}_1$ and $\mathcal{K}_2$ are subspaces of $\mathbb{R}^{n+m}$, then $\mathcal{K}_1^{\perp} \cap \mathcal{K}_2^{\perp} = \{\mathbf{0}\}$ if and only if $\mathcal{K}_1 + \mathcal{K}_2 = \mathbb{R}^{n+m}$.*

The following lemma will be needed in the proof of forthcoming Lemma 5.17.

Lemma 5.12. *If the sets $\mathbb{B}(\mathbf{0}, 1)$ and D are homeomorphic, then any homeomorphism f between $\mathbb{S}(\mathbf{0}, 1)$ and ∂D can be extended to a homeomorphism $F : \mathbb{B}(\mathbf{0}, 1) \to D$ so that $F\,|_{\mathbb{S}(\mathbf{0},1)} = f$.*

Proof. Let $H : \mathbb{B}(\mathbf{0}, 1) \to D$ be a homeomorphism between $\mathbb{B}(\mathbf{0}, 1)$ and D and denote $h : \mathbb{S}(\mathbf{0}, 1) \to \partial D$ by $h = H\,|_{\mathbb{S}(\mathbf{0},1)}$. We define the (continuous) function $F : \mathbb{B}(\mathbf{0}, 1) \to D$ by

$$F(x) = \begin{cases} H\left(|x|\, h^{-1} f\left(x/|x|\right)\right) & \text{if } x \neq \mathbf{0} \\ \\ H(\mathbf{0}) & \text{if } x = \mathbf{0}. \end{cases}$$

It is straightforward to check that $F\,|_{\mathbb{S}(\mathbf{0},1)} = f$. Let us show that F is injective: indeed, if $F(x_1) = F(x_2)$, then $|x_1|\, h^{-1}(f\left(x_1/|x_1|\right)) = |x_2|\, h^{-1}(f\left(x_2/|x_2|\right))$. If both sides are zero, then $x_1 = x_2 = \mathbf{0}$. Otherwise $|x_1| = |x_2|$ and $x_1/|x_1| = x_2/|x_2|$, which implies that $x_1 = x_2$.

To see that F is a bijection, fix any $y \in D$, and let $x' \in \mathbb{B}(\mathbf{0}, 1)$ be such that $y = H(x')$. If $x' = \mathbf{0}$, then $y = F(\mathbf{0})$. Otherwise,

$$y = H\left(|x'|\left(\frac{x'}{|x'|}\right)\right) = H(|x'|\; h^{-1} \circ f\left(f^{-1} \circ h(\frac{x'}{|x'|})\right) = F\left(|x'|\, f^{-1} \circ h(\frac{x'}{|x'|})\right).$$

This shows that F is also surjective, thus a continuous bijection. Since $\mathbb{B}(\mathbf{0}, 1)$ is compact, it follows that F is a homeomorphism. □

We end this section by introducing the notion of *linking* that is commonly used in critical point theory. Let us fix some terminology: if $B \subset \mathbb{R}^n$ is homeomorphic

to a subset of $\mathbb{R}^d$ with nonempty interior, we say that the set ∂B is the *relative boundary* of B if it is a homeomorphic image of the boundary of the set in $\mathbb{R}^d$.

Definition 5.13. [82, Section II.8] Let A be a subset of $\mathbb{R}^{n+m}$ and let B be a submanifold of $\mathbb{R}^{n+m}$ with relative boundary ∂B. Then we say that A and $\Gamma = \partial B$ *link* if

(i) $A \cap \Gamma = \emptyset$

(ii) for any continuous map $h \in \mathcal{C}^0\left(\mathbb{R}^{n+m}, \mathbb{R}^{n+m}\right)$ such that $h \mid_\Gamma = id$ we have $h(B) \cap A \neq \emptyset$.

In particular, the following result holds.

Theorem 5.14. *[Linking sets] Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be linear subspaces such that $\mathcal{K}_1 \oplus \mathcal{K}_2 = \mathbb{R}^{n+m}$, and take any $\bar{v} \in \mathcal{K}_1 \backslash \{\mathbf{0}\}$. Then for $0 < r < R$, the sets*

$$A := \mathbb{S}(\mathbf{0}, r) \cap \mathcal{K}_1 \qquad \textit{and} \qquad \Gamma := \left(\mathbb{B}(\mathbf{0}, R) \cap \mathcal{K}_2\right) \cup \left(\mathbb{S}(\mathbf{0}, R) \cap \left(\mathcal{K}_2 + \mathbb{R}_+ \{\bar{v}\}\right)\right)$$

link.

Proof. Use methods in [82, Section II.8], or infer from Example 3 there. □

5.4 More on the structure of semi-algebraic maps

In the sequel we shall always consider a set-valued map $S : \mathcal{X} \rightrightarrows \mathbb{R}^m$, where $\mathcal{X} \subset \mathbb{R}^n$, and we shall assume that S is semi-algebraic.

Theorem 5.15. *Assume that $S : \mathcal{X} \rightrightarrows \mathbb{R}^m$ is outer semicontinuous, and the sets $\mathcal{X} \subset \mathbb{R}^n$ and $\operatorname{gph}(S)$ are semi-algebraic. Then S is strictly continuous with respect to $\mathcal{X}$ everywhere except on a set of dimension at most $(\dim \mathcal{X} - 1)$.*

Proof. Using Theorem 2.17 we stratify $\mathcal{X}$ into a disjoint union of manifolds (strata) $\{\mathcal{X}_j\}_j$ and study how S behaves on the strata $\mathcal{X}_j$ of full dimension (that is, $\dim(\mathcal{X}_j) = \dim(\mathcal{X}) = d \leq n$). For each such stratum $\mathcal{X}_j$, if S is not strictly continuous at $\bar{x} \in \mathcal{X}_j$ relative to $\mathcal{X}_j$, then by [80, Theorem 9.38], there is some $\bar{y} \in S(\bar{x})$ such that $\operatorname{lip}_{\mathcal{X}_j} S(\bar{x} \mid \bar{y}) = \infty$. Since S is outer semicontinuous, we deduce from Proposition 5.9 that there is a nonzero vector $v \in N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \cap (T_{\mathcal{X}_j}(\bar{x}) \times \{\mathbf{0}_m\})$.

We now stratify the semi-algebraic set $\operatorname{gph}(S) \cap (\mathcal{X}_j \times \mathbb{R}^m)$ into a finite union of disjoint manifolds $\{\mathcal{M}_k\}_k$. Since $v \in N_{\operatorname{gph}(S)}(\bar{x}, \bar{y}) \setminus \{\mathbf{0}_{n+m}\}$, it can be written as a limit of Hadamard normal vectors $v_i \in \hat{N}_{\operatorname{gph}(S)}(x_i, y_i)$ with $(x_i, y_i) \to (\bar{x}, \bar{y})$. Passing to a subsequence if necessary, we may assume that the sequence $\{(x_i, y_i)\}_i$ belongs to the same stratum, say $\mathcal{M}_{k^*}$ and $v_i \in \hat{N}_{\mathcal{M}_{k^*}}(x_i, y_i)$ (note that $\mathcal{M}_{k^*} \subset \operatorname{gph}(S)$). Since $\mathcal{M}_{k^*}$ is a smooth manifold, we have $\hat{N}_{\mathcal{M}_{k^*}}(x_i, y_i) = N_{\mathcal{M}_{k^*}}(x_i, y_i) = [T_{\mathcal{M}_{k^*}}(x_i, y_i)]^{\perp}$. Using the Whitney property (normal regularity) of the stratification, we deduce that v must lie in some $N_{\mathcal{M}}(\bar{x}, \bar{y}) \cap \big(T_{\mathcal{X}_j}(\bar{x}) \times \{\mathbf{0}_m\}\big)$, where $\mathcal{M}$ is the stratum that contains $(\bar{x}, \bar{y})$. Lemma 5.10 then tells us that the set of all possible $\bar{x}$ is of lower dimension than that of $\mathcal{X}_j$ (or $\mathcal{X}$). Since there are finitely many strata $\mathcal{X}_j$, the result follows. □

Remark. Note that the domain of S

$$\operatorname{dom}(S) := \{x \in \mathcal{X} : S(x) \neq \emptyset\},$$

being the projection to $\mathbb{R}^n$ of the semi-algebraic set $\operatorname{gph}(S)$, is always semi-algebraic. Thus, if S has nonempty values, the above assumption "$\mathcal{X}$ semi-algebraic" becomes superfluous. In any case, one can eliminate this assumption from the statement and replace $\mathcal{X}$ by $\mathcal{X}' := \operatorname{dom}(S)$ the domain of S.

The next lemma will be crucial in the sequel. We shall first need some notation.

In the sequel we denote by

$$\mathcal{L} := \{\mathbf{0}_n\} \times \mathbb{R}^m \tag{5.4.1}$$

as a subspace of $\mathbb{R}^n \times \mathbb{R}^m$ and we denote by $\bar{S} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ the set-valued map whose graph is the closure of the graph of S, that is,

$$\mathrm{gph}\,(\bar{S}) = \mathrm{cl}(\mathrm{gph}\,(S)).$$

Lemma 5.16. *Let $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a closed-valued semi-algebraic set-valued map. For any $k > 0$, there is a $\mathcal{C}^k$ stratification $\{\mathcal{M}_i\}_i$ of $\mathrm{gph}\,(S)$ such that if $S(\bar{x}) \neq \bar{S}(\bar{x})$ for some $\bar{x} \in \mathbb{R}^n$, then there exist $\bar{y} \in \mathbb{R}^m$, a stratum $\mathcal{M}_i$ of the stratification of $\mathrm{gph}\,(S)$ and a neighborhood U of $(\bar{x}, \bar{y})$ such that $(\bar{x}, \bar{y}) \in \mathrm{cl}\,(\mathcal{M}_i)$ and*

$$((\bar{x}, \bar{y}) + \mathcal{L}) \cap \mathcal{M}_i \cap U = \emptyset.$$

Proof. By Theorem 2.17 we stratify $\mathrm{gph}\,(S)$ into a disjoint union of finitely many manifolds, that is $\mathrm{gph}\,(S) = \cup_i \mathcal{M}_i$. Consider the set-valued map $S_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ whose graph consists of the manifold $\mathcal{M}_i$. Let further $\dot{S}_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be the map such that $\dot{S}_i(x) = \mathrm{cl}\,(S_i(x))$ for all x, and $\bar{S}_i : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be the map whose graph is $\mathrm{cl}(\mathrm{gph}\,(S_i))$, also equal to $\mathrm{cl}(\mathrm{gph}\,(\dot{S}_i))$. Both $\dot{S}_i$ and $\bar{S}_i$ are semi-algebraic (for example, [33]), and there exists a stratification of $\mathrm{cl}(\mathrm{gph}\,(S))$ such that the graphs of S_i, $\dot{S}_i$ and $\bar{S}_i$ can be represented as a finite union of strata of that stratification, by Theorem 2.17 again.

We now prove that if $S(\bar{x}) \neq \bar{S}(\bar{x})$, then there is some i such that $\dot{S}_i$ is not outer semicontinuous at $\bar{x}$. Indeed, in this case there exists $\bar{y}$ such that $(\bar{x}, \bar{y}) \in \mathrm{cl}(\mathrm{gph}\,(S))\backslash\mathrm{gph}\,(S)$. Note that $\mathrm{cl}(\mathrm{gph}\,(S)) = \cup_i \mathrm{gph}\,(\bar{S}_i)$. This means that $(\bar{x}, \bar{y})$ must lie in $\mathrm{gph}\,(\bar{S}_i)\backslash\mathrm{gph}\,(\dot{S}_i)$ for some i, which means that $\dot{S}_i$ is not outer semicontinuous at $\bar{x}$ as claimed.

Obviously $(\bar{x},\bar{y}) \in \mathrm{cl}(\mathcal{M}_i)$. Suppose that $((\bar{x},\bar{y})+\mathcal{L})\cap\mathcal{M}_i\cap U \neq \emptyset$ for all neighborhoods U containing $(\bar{x},\bar{y})$. Then there is a sequence $y_j \to \bar{y}$ such that $(\bar{x},y_j) \in \mathcal{M}_i$. Since $\dot{S}_i$ is closed-valued, this would yield $(\bar{x},\bar{y}) \in \mathrm{gph}\,(\dot{S}_i)$, which contradicts $(\bar{x},\bar{y}) \notin \mathrm{gph}\,(\dot{S}_i)$ earlier. □

Keeping now the notation of the proof of the previous lemma, let us set $\bar{z} := (\bar{x},\bar{y})$. Let further $\mathcal{M}_i, \mathcal{M}'$ be the strata of $\mathrm{cl}(\mathrm{gph}\,(S))$ such that $z \in \mathcal{M}' \subset \mathrm{cl}(\mathcal{M}_i)$. In the next lemma we are working with normals on manifolds, so it does not matter which kind of normal in Definition 2.7 we consider.

Lemma 5.17. *Suppose there is a neighborhood U of $\bar{z}$ such that $\bar{z} \in \mathcal{M}'$, $\mathcal{M}' \subset \mathrm{cl}(\mathcal{M}_i)$ and $(\bar{z}+\mathcal{L})\cap\mathcal{M}_i\cap U = \emptyset$, where $\mathcal{L}$ is defined in* (5.4.1)*. Then $N_{\mathcal{M}'}(\bar{z})\cap\mathcal{L}^{\perp} \supsetneq \{\mathbf{0}_{n+m}\}$.*

Proof. We prove the result by contradiction. Suppose that $N_{\mathcal{M}'}(\bar{z})\cap\mathcal{L}^{\perp} = \{\mathbf{0}_{n+m}\}$. Then $T_{\mathcal{M}'}(\bar{z}) + \mathcal{L} = \mathbb{R}^{n+m}$ by Proposition 5.11. We may assume, by taking a submanifold of $\mathcal{M}'$ if necessary, that $\dim\mathcal{M}' = n$ so that $\dim\mathcal{M}' + \dim\mathcal{L} = n+m$ and $T_{\mathcal{M}'}(\bar{z}) \oplus \mathcal{L} = \mathbb{R}^{n+m}$. Owing to the so-called wink lemma (see [39, Proposition 5.10] *e.g.*) we may assume that $\dim\mathcal{M}_i = n+1$.

(Case $m = 1$) We first consider the case where $m = 1$. In this case, the subspace $\mathcal{L}$ is a line whose spanning vector $v = (\mathbf{0},1)$ is not in $T_{\mathcal{M}'}(\bar{z})$. There is a neighborhood U' of $\bar{z}$ such that $U' \subset U$, $\mathcal{M}'\cap U'$ equals $f^{-1}(0)$ for some smooth function $f : U' \to \mathbb{R}$ (local equation of $\mathcal{M}'$), and $\mathcal{M}_i \cap U' = f^{-1}((0,\infty))$. The gradient $\nabla f(\bar{z})$ is nonzero and is not orthogonal to v since $T_{\mathcal{M}'}(\bar{z})$ is the set of vectors orthogonal to $\nabla f(\bar{z})$ and $T_{\mathcal{M}'}(\bar{z}) \oplus \mathcal{L} = \mathbb{R}^{n+1}$. There are points in $(\bar{z}+\mathcal{L})\cap U'$ such that f is positive, which means that $(\bar{z}+\mathcal{L})\cap\mathcal{M}_i\cap U' \neq \emptyset$, contradicting the stipulated conditions. Therefore, we assume that $m > 1$ for the

rest of the proof.

(Case $m > 1$) As in the previous case, we shall eventually prove that $(\bar{z} + \mathcal{L}) \cap \mathcal{M}_i \cap U' \neq \emptyset$ reaching to a contradiction. To this end, let us denote by h_0 the (semi-algebraic) homeomorphism of $\mathbb{R}^{n+m}$ to $\mathbb{R}^{n+m}$ which, for some neighborhood $V \subset U$ of $\bar{z}$, maps homeomorphically $V \cap (\mathcal{M}_i \cup \mathcal{M}')$ to $\mathbb{R}^n \times (\mathbb{R}_+ \times \{\mathbf{0}_{m-1}\}) \subset \mathbb{R}^{n+m}$ and $V \cap \mathcal{M}'$ to $\mathbb{R}^n \times \{\mathbf{0}_m\}$ (see [33, Theorem 3.12] *e.g.*).

Claim. We first show that there exists a closed neighborhood $W \subset V$ of $\bar{z}$ such that $W \cap \mathcal{M}'$ and $\partial W \cap \mathcal{M}_i$ are both homeomorphic to $\mathbb{B}^n$ and $W \cap \mathcal{M}' = \mathbb{B}^{n+m}(\bar{z}, R_1) \cap \mathcal{M}'$ for some $R_1 > 0$.

Since $\mathcal{M}'$ is a smooth manifold, there exists $R_1 > 0$ such that $\mathbb{B}^{n+m}(\bar{z}, R_1) \cap \mathcal{M}'$ is homeomorphic (in fact, diffeomorphic) to $(T_{\mathcal{M}'}(\bar{z}) + \bar{z}) \cap \mathbb{B}^{n+m}(\bar{z}, R_1)$, which in turn is homeomorphic to $\mathbb{B}^n$, as is shown by the homeomorphism:

$$z \mapsto \left(\frac{|z - \bar{z}|}{|P(z) - \bar{z}|} (P(z) - \bar{z}) \right) + \bar{z}\,,$$

where P denotes the projection onto the tangent space $\bar{z} + T_{\mathcal{M}'}(\bar{z})$. Consider the image of $\mathbb{B}^{n+m}(\bar{z}, R_1) \cap \mathcal{M}'$ under the map h_0. This image lies in the set $\mathbb{R}^n \times \{\mathbf{0}_m\}$. Therefore, for $r_1 > 0$ sufficiently small, the set $W = h_0^{-1}\left(h_0\left(\mathbb{B}^{n+m}(\bar{z}, R_1) \cap \mathcal{M}'\right) + [-r_1, r_1]^m\right)$ satisfies the required properties, concluding the proof of our claim.

Let us further fix $v \in \mathcal{L} \setminus \{\mathbf{0}\}$ and consider the set

$$\Gamma' := \underbrace{\left(\mathbb{B}^{n+m}(\bar{z}, R) \cap (\bar{z} + T_{\mathcal{M}'}(\bar{z}))\right)}_{\Gamma'_1} \cup \underbrace{\left(\mathbb{S}^{n+m-1}(\bar{z}, R) \cap (\bar{z} + T_{\mathcal{M}'}(\bar{z}) + \mathbb{R}_+ \{v\})\right)}_{\Gamma'_2}\,.$$

Setting

$$A := \mathbb{S}^{n+m-1}(\bar{z}, r) \cap \mathcal{L}, \text{ where } 0 < r < R\,,$$

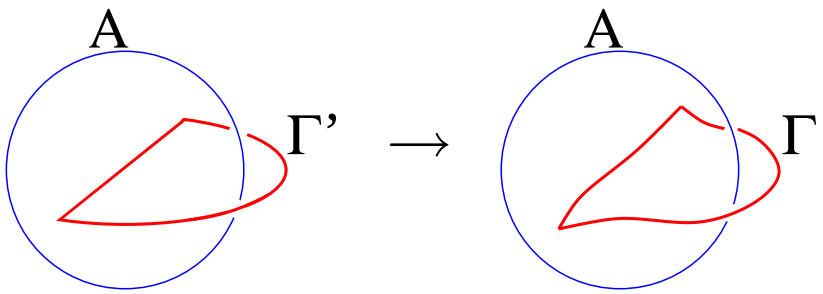


Figure 5.1: Linking sets (A, Γ') and (A, Γ).

we immediately get that the sets A and Γ' link (*c.f.* Theorem 5.14). Based on this, our objective is to prove that the sets A and Γ also link, where Γ is defined by

$$\Gamma = \underbrace{(W \cap \mathcal{M}')}_{\Gamma_1} \cup \underbrace{(\partial W \cap (\mathcal{M}_i \cup \mathcal{M}'))}_{\Gamma_2},$$

provided $r > 0$ is chosen appropriately. Once we succeed in doing so, we apply Definition 5.13 (for $h = id$) to deduce that $(\bar{z} + \mathcal{L}) \cap \mathcal{M}_i \cap U \neq \emptyset$, which contradicts our initial assumptions. Figure 5.1 illustrates the sets A, Γ and Γ' for $n = 1$ and $m = 2$.

For the sequel, we introduce the notation "$\overset{\simeq}{\rightarrow}$" in $f : D_1 \overset{\simeq}{\rightarrow} D_2$ to mean that f is a homeomorphism between the sets D_1 and D_2. In Step 1 and Step 2, we define a continuous function $H : (\mathbb{B}^{n+1}(\mathbf{0}, 1) \times \{0\}) \cup (\mathbb{S}^n(\mathbf{0}, 1) \times [0, 2]) \rightarrow \mathbb{B}^{n+m}(\bar{z}, R)$ that will be used in Step 3.

Step 1: Determine H on $(\mathbb{B}^{n+1}(\mathbf{0}, 1) \times \{0\}) \cup (\mathbb{S}^n(\mathbf{0}, 1) \times [0, 2])$.

In Steps 1 (a) to 1 (c), we define a continuous function H on $\mathbb{S}^n(\mathbf{0}, 1) \times [0, 2]$ so that $H\,|_{\mathbb{S}^n(\mathbf{0},1)\times[0,2]}$ is a homotopy between Γ and Γ'. More precisely, denoting by

$$\mathbb{S}^n_+(\mathbf{0}, 1) := \mathbb{S}^n(\mathbf{0}, 1) \cap (\mathbb{R}^n \times [0, \infty)),$$
$$\mathbb{S}^n_-(\mathbf{0}, 1) := \mathbb{S}^n(\mathbf{0}, 1) \cap (\mathbb{R}^n \times (-\infty, 0]),$$

we want to define H in such a way that its restrictions

$$H(\cdot,0)\mid_{\mathbb{S}^n_+(\mathbf{0},1)}\colon \mathbb{S}^n_+(\mathbf{0},1) \xrightarrow{\simeq} \Gamma_1 \subset \mathbb{R}^{n+m},$$
$$H(\cdot,0)\mid_{\mathbb{S}^n_-(\mathbf{0},1)}\colon \mathbb{S}^n_-(\mathbf{0},1) \xrightarrow{\simeq} \Gamma_2 \subset \mathbb{R}^{n+m},$$
$$H(\cdot,2)\mid_{\mathbb{S}^n_+(\mathbf{0},1)}\colon \mathbb{S}^n_+(\mathbf{0},1) \xrightarrow{\simeq} \Gamma'_1 \subset \mathbb{R}^{n+m},$$
$$H(\cdot,2)\mid_{\mathbb{S}^n_-(\mathbf{0},1)}\colon \mathbb{S}^n_-(\mathbf{0},1) \xrightarrow{\simeq} \Gamma'_2 \subset \mathbb{R}^{n+m},$$

are homeomorphisms between the respective spaces. Note that both $\mathbb{S}^n_+(\mathbf{0},1)$ and $\mathbb{S}^n_-(\mathbf{0},1)$ are homeomorphic to $\mathbb{B}^n(\mathbf{0},1)$. For notational convenience, we denote by $\mathbb{S}^n_=(\mathbf{0},1)$ the set $\mathbb{S}^n(\mathbf{0},1)\cap(\mathbb{R}^n\times\{0\}) = \mathbb{S}^{n-1}(\mathbf{0},1)\times\{0\}$.

Step 1 (a). **Determine H on $\mathbb{S}(\mathbf{0},1)\times[0,1]$.**
Since $\partial W \cap \operatorname{cl}\mathcal{M}_i$ is a closed set that does not contain $\bar{z}$, there is some $R>0$ such that $(\partial W\cap\mathcal{M}_i)\cap\mathbb{B}^{n+m}(\bar{z},R)=\emptyset$ and $\mathbb{B}^{n+m}(\bar{z},R)\subset U$. We proceed to create the homotopy H so that

$$H(\cdot,1)\mid_{\mathbb{S}^n_+(\mathbf{0},1)}\colon \mathbb{S}^n_+(\mathbf{0},1) \xrightarrow{\simeq} \Gamma''_1 \subset \mathbb{R}^{n+m},$$
$$H(\cdot,1)\mid_{\mathbb{S}^n_-(\mathbf{0},1)}\colon \mathbb{S}^n_-(\mathbf{0},1) \xrightarrow{\simeq} \Gamma''_2 \subset \mathbb{R}^{n+m},$$

where

$$\Gamma''_1 = \mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}',$$
$$\text{and}\qquad \Gamma''_2 \subset \mathbb{S}^{n+m-1}(\bar{z},R) \text{ is homeomorphic to } \Gamma_2.$$

The first homotopy between Γ_1 and Γ''_1 can be chosen such that $H(s,t)\in\mathcal{M}'$ for all $s\in\mathbb{S}^n_+(\mathbf{0},1)$ and $t\in[0,1]$. We also require that $d(\bar{z},H(s,t))\geq R$ for all $s\in\mathbb{S}^n_=(\mathbf{0},1)$ and $t\in[0,1]$, which does not present any difficulties.

For the second homotopy between Γ_2 and Γ''_2, we first extend $H(\cdot,1)$ so that $H(\cdot,1)\mid_{\mathbb{S}^n(\mathbf{0},1)}\colon \mathbb{S}^n(\mathbf{0},1)\xrightarrow{\simeq}\Gamma''_1\cup\Gamma''_2$ is a homeomorphism between the corresponding spaces. This is achieved by showing that there is a homeomorphism $H(\cdot,1)\mid_{\mathbb{S}^n_-(\mathbf{0},1)}$

between $\mathbb{S}^n_-(\mathbf{0},1)$ and Γ_2''. Let $h_2 : \mathbb{B}^n(\mathbf{0},1) \xrightarrow{\simeq} \mathbb{S}^n_-(\mathbf{0},1)$ be a homeomorphism between $\mathbb{B}^n(\mathbf{0},1)$ and $\mathbb{S}^n_-(\mathbf{0},1)$. Then $H(\cdot,1)\mid_{\mathbb{S}^n_=(\mathbf{0},1)} \circ h_2 \mid_{\mathbb{S}^{n-1}(\mathbf{0},1)}: \mathbb{S}^{n-1}(\mathbf{0},1) \xrightarrow{\simeq} \partial\Gamma_2''$. By Lemma 5.12 this can be extended to a homeomorphism $G : \mathbb{B}^n(\mathbf{0},1) \xrightarrow{\simeq} \Gamma_2''$. Define $H(\cdot,1)\mid_{\mathbb{S}^n_-(\mathbf{0},1)}: \mathbb{S}^n_-(\mathbf{0},1) \xrightarrow{\simeq} \Gamma_2''$ by $H(\cdot,1)\mid_{\mathbb{S}^n_-(\mathbf{0},1)} = G \circ h_2^{-1}$.

It remains to resolve H on $\mathbb{S}^n_-(\mathbf{0},1)\times(0,1)$. Note that the sets

$$H(\mathbb{S}^n_=(\mathbf{0},1)\times[0,1]), \quad H(\mathbb{S}^n_-(\mathbf{0},1)\times\{0\}) = \Gamma_2 \quad \text{and} \quad H(\mathbb{S}^n_-(\mathbf{0},1)\times\{1\}) = \Gamma_2''$$

are all of dimension at most n, so the radial projection of these sets onto $\mathbb{S}^{n+m-1}(\bar{z},R)$ is of dimension at most n. Since $\mathbb{S}^{n+m-1}(\bar{z},R)$ is of dimension at least $n+1$, we can find some point $p \in \mathbb{S}^{n+m-1}(\bar{z},R)$ not lying in the radial projections of these sets. The set

$$D := \mathbb{R}^{n+m}\backslash(((\mathbb{R}_+\{p-\bar{z}\})+\{\bar{z}\}) \cup \mathbb{B}^{n+m}(\bar{z},R))$$

is homeomorphic to $\mathbb{R}^{n+m}$, so by the Tietze extension theorem (see for example [74]), we can extend H continuously to $\mathbb{S}^n_-(\mathbf{0},1)\times[0,1]$ so that $H(\mathbb{S}^n_-(\mathbf{0},1)\times[0,1]) \subset D$.

Step 1 (b). **Determine H on $\mathbb{S}^n_+(\mathbf{0},1)\times[1,2]$.** We next define $H\mid_{\mathbb{S}^n_+(\mathbf{0},1)\times[1,2]}$, the homotopy between Γ_1'' and Γ_1'. Since $\mathcal{M}'$ is a manifold, for any $\delta > 0$, we can find R small enough such that for any $z \in \mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}'$, the distance from z to $\bar{z}+T_{\mathcal{M}'}(\bar{z})$ is at most δR. The value R can be reduced if necessary so that the mapping P, which projects a point $z \in \mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}'$ to the closest point in $\bar{z}+T_{\mathcal{M}'}(\bar{z})$, is a homeomorphism of $\mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}'$ to its image.

Define the map $H_1 : (\mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}')\times[1,2] \to \mathbb{B}^{n+m}(\bar{z},R)$ by

$$H_1(z,t) := \left(\frac{|z-\bar{z}|}{|(2-t)z+(t-1)P(z)-\bar{z}|}\left((2-t)z+(t-1)P(z)-\bar{z}\right)\right)+\bar{z}.$$

This is a homotopy from Γ_1 to Γ_1'. For any homeomorphism $h_1 : \mathbb{B}^{n+m}(\bar{z},R)\cap\mathcal{M}' \xrightarrow{\simeq} \mathbb{S}^n_+(\mathbf{0},1)$, we define $H\mid_{\mathbb{S}^n_+(\mathbf{0},1)\times[0,1]}$ via $H(s,t) = H_1\left(h_1^{-1}(s),t\right)$.

Step 1 (c). **Determine H on $\mathbb{S}^n_-(\mathbf{0},1)\times[1,2]$.** We now define $H\mid_{\mathbb{S}^n_-(\mathbf{0},1)\times[1,2]}$, the homotopy between Γ_2'' and Γ_2' that respects the boundary conditions stipulated by $H\mid_{\mathbb{S}^n_=(\mathbf{0},1)\times[1,2]}$. We extend $H(\cdot,1)\mid_{\mathbb{S}^n(\mathbf{0},1)}$ so that it is a homeomorphism between $\mathbb{S}^n(\mathbf{0},1)$ and $\Gamma_1'\cup\Gamma_2'$ by using methods similar to that used in Step $1(a)$.

We now use the Tietze extension theorem to establish a continuous extension of H to $\mathbb{S}^n(\mathbf{0},1)\times[1,2]$. We are left only to resolve H on $\mathbb{S}^n_-(\mathbf{0},1)\times(1,2)$. Much of this is now similar to the end of step $1(a)$. The dimension of $\mathbb{S}^{n+m-1}(\bar{z},R)$ is $n+m-1$, while the dimensions of Γ_2'', Γ_2' and $H\left(\mathbb{S}^n_=(\mathbf{0},1)\times[1,2]\right)$ are all at most n. Therefore, there is one point in $\mathbb{S}^{n+m-1}(\bar{z},R)$ outside these three sets, say p. Since $\mathbb{S}^{n+m-1}(\bar{z},R)\backslash\{p\}$ is homeomorphic to $\mathbb{R}^{n+m-1}$, the Tietze extension theorem again implies that we can extend H continuously in $\mathbb{S}^n(\mathbf{0},1)\times[1,2]$.

Step 1 (d). **Determine H on $\mathbb{B}^{n+1}(\mathbf{0},1)\times\{0\}$.** We use Lemma 5.12 to extend the domain of the function

$$H(\cdot,0):\mathbb{S}^n(\mathbf{0},1)\xrightarrow{\simeq}(\mathcal{M}'\cap W)\cup(\mathcal{M}_i\cap\partial W)$$

to $\mathbb{B}^{n+1}(\mathbf{0},1)$ so that

$$H(\cdot,0):\mathbb{B}^{n+1}(\mathbf{0},1)\xrightarrow{\simeq}(\mathcal{M}'\cup\mathcal{M}_i)\cap W$$

is a homeomorphism.

Step 2: Choice of R and r. We now choose R and r so that $H\left(\mathbb{S}^n(\mathbf{0},1)\times[0,2]\right)$ does not intersect $A=\mathbb{S}^{n+m-1}(\bar{z},r)\cap(\bar{z}+\mathcal{L})$. To this end, consider the minimization problem

$$\min\left\{\text{dist}\,(z,T_{\mathcal{M}'}(\bar{z})+\bar{z}):z\in\mathbb{S}^{n+m-1}(\bar{z},r)\cap(\bar{z}+\mathcal{L})\right\}.$$

Since $\mathbb{S}^{n+m-1}(\bar{z},r)\cap(\bar{z}+\mathcal{L})$ is compact, the above minimum is attained at some point z_r and its value is not zero (otherwise $z_r-\bar{z}$ would be a nonzero element

in $T_{\mathcal{M}'}(\bar{z}) \cap \mathcal{L}$, contradicting $T_{\mathcal{M}'}(\bar{z}) \cap \mathcal{L} = \{\mathbf{0}\}$). Therefore, for some constant $\varepsilon \in (0,1]$ independent of r, it holds dist $(z_r, T_{\mathcal{M}'}(\bar{z}) + \bar{z}) = \varepsilon\, r$.

Given $\delta > 0$, we can shrink R if necessary to get $\mathrm{d}\,(z, T_{\mathcal{M}'}(\bar{z}) + \bar{z}) \leq \delta\, R$ for all $z \in H\left(\mathbb{S}^n_+(\mathbf{0},1) \times [0,1]\right)$. If $\delta < \varepsilon$, we can find some r satisfying $\delta R < \varepsilon\, r \leq r < R$. Since $\delta R < \varepsilon\, r$, $H\left(\mathbb{S}^n_+(\mathbf{0},1) \times [1,2]\right)$ does not intersect $\mathbb{S}^{n+m-1}(\bar{z}, r) \cap (\bar{z} + \mathcal{L})$. From $r < R$, it is clear that $H\left(\mathbb{S}^n_-(\mathbf{0},1) \times [0,2]\right)$, being a subset of $\mathrm{cl}\,(\mathbb{R}^{n+m} \backslash \mathbb{B}^{n+m}(\bar{z}, R))$, does not intersect $\mathbb{S}^{n+m-1}(\bar{z}, r) \cap (\bar{z} + \mathcal{L})$. Elements in $H\left(\mathbb{S}^n_+(\mathbf{0},1) \times [0,1]\right)$ are either in $\mathbb{B}^{n+m}(\bar{z}, R) \cap \mathcal{M}'$ or outside $\mathbb{B}^{n+m}(\bar{z}, R)$, so $H\left(\mathbb{S}^n(\mathbf{0},1) \times [0,2]\right)$ does not intersect A as needed.

Step 3: Set-up for linking theorem. Let

$$h_3 : \mathbb{B}^{n+1}(\mathbf{0},1) \xrightarrow{\simeq} \left(\mathbb{B}^{n+1}(\mathbf{0},1) \times \{0\}\right) \cup \left(\mathbb{S}^n(\mathbf{0},1) \times [0,2]\right)$$

be a homeomorphism between the respective spaces. We can extend the homeomorphism

$$H\,|_{\mathbb{S}^n(\mathbf{0},1)\times\{2\}} \circ h_3\,|_{\mathbb{S}^n(\mathbf{0},1)} \colon \mathbb{S}^n(\mathbf{0},1) \xrightarrow{\simeq} \Gamma'$$

to

$$h_4 : \mathbb{B}^{n+1}(\mathbf{0},1) \xrightarrow{\simeq} (T_{\mathcal{M}'}(\bar{z}) + \mathbb{R}_+\,\{v\} + \bar{z}) \cap \mathbb{B}^{n+m}(\bar{z}, R).$$

Define the map

$$g : (T_{\mathcal{M}'}(\bar{z}) + \mathbb{R}_+\,\{v\} + \bar{z}) \cap \mathbb{B}^{n+m}(\bar{z}, R) \to \mathbb{B}^{n+m}(\bar{z}, R)$$

by $g = H \circ h_3 \circ h_4^{-1}$. By construction, the map $g\,|_{\Gamma'}$ is the identity map there. Furthermore, g can be extended continuously to the domain $\mathbb{R}^{n+m}$ by the Tietze extension theorem.

Step 4: Apply linking theorem. Recall that $A := \mathbb{B}^{n+m}(\bar{z}, r) \cap (\bar{z} + \mathcal{L})$ and Γ' link by Theorem 5.14. This means that there is a nonempty intersection of

$g\left((T_{\mathcal{M}'}(\bar{z}) + \mathbb{R}_+\{v\} + \bar{z}) \cap \mathbb{B}^{n+m}(\bar{z}, R)\right)$ with A. Step 2 asserts that the intersection is not in $H\left(\mathbb{S}^n(\mathbf{0}, 1) \times [0, 2]\right)$, so the intersection lies in $H(\mathbb{B}^{n+1}(\mathbf{0}, 1) \times \{0\})$. In other words, A and Γ link. This means that $W \cap \mathcal{M}_i$ intersects $\mathbb{B}^{n+m}(\bar{z}, r) \cap (\bar{z} + \mathcal{L})$, which means that $(\bar{z} + \mathcal{L}) \cap \mathcal{M}_i \cap U \neq \emptyset$, contradicting our assumption. □

5.5 Main result

In this section we put together all previous results to obtain the following theorem. Recall that $\bar{S}$ is the set-valued map whose graph is the closure of the graph of S (thus, $\bar{S}$ is outer semicontinuous by definition).

Theorem 5.18. *If $S : \mathcal{X} \rightrightarrows \mathbb{R}^m$ is a closed-valued semi-algebraic set-valued map, where $\mathcal{X} \subset \mathbb{R}^n$ is semi-algebraic, then S and $\bar{S}$ differ outside a set of dimension at most $(\dim \mathcal{X} - 1)$.*

Proof. We first consider the case where $\mathcal{X} = \mathbb{R}^n$ and a $\mathcal{C}^k$ stratification of $\mathrm{cl}(\mathrm{gph}\,(S))$. If $S(\bar{x}) \neq \bar{S}(\bar{x})$, then Lemma 5.16 and Lemma 5.17 yield that there exists some $\bar{y}$ and stratum $\mathcal{M}'$ containing $\bar{z} := (\bar{x}, \bar{y})$ such that $N_{\mathcal{M}'}(\bar{z}) \cap \mathcal{L}^\perp \supsetneq \{\mathbf{0}_{n+m}\}$. Finally, since there are only finitely many strata, Lemma 5.10 tells us that $S(x)$ and $\bar{S}(x)$ may differ only on a set of dimension at most $n - 1$. This proves the result in this particular case.

We now consider the case where $\mathcal{X} \neq \mathbb{R}^n$. Let $\mathcal{X} = \dot{\cup}\mathcal{X}_j$ be a stratification of $\mathcal{X}$, and let $\mathcal{D}$ be the union of strata of full dimension in $\mathcal{X}$. Each stratum in $\mathcal{D}$ is semi-algebraically homeomorphic to $\mathbb{R}^d$, where $d := \dim \mathcal{X}$ and let $h_j : \mathbb{R}^d \to \mathcal{X}_j$ denote such a homeomorphism. By considering the set-valued maps $S \circ h_j$ for all j, we reduce the problem to the aforementioned case. Since the set of strata (*a*

fortiori the set of full-dimensional strata) is finite, we deduce that $S(x) \neq \bar{S}(x)$ can only occur in a set of dimension at most $d-1$. □

The following result is now an easy consequence of the above.

Theorem 5.19. *[Main result] A closed-valued semi-algebraic set-valued map* $S : \mathcal{X} \rightrightarrows \mathbb{R}^m$*, where* $\mathcal{X} \subset \mathbb{R}^n$ *is semi-algebraic, is strictly continuous outside a set of dimension at most* $(\dim \mathcal{X} - 1)$.

Proof. By Theorem 5.18 the map S differs from the outer semicontinuous map $\bar{S}$ on a set of dimension at most $(\dim \mathcal{X} - 1)$. Apply Theorem 5.15. □

Remark. Our main result (Theorem 5.19) as well as all previous preliminary results (Lemmas 5.16, 5.17, Theorems 5.15, 5.18) can be restated for the case where S is definable in an o-minimal structure. With slightly more effort we can further extend these results in case where S is tame, noting that one performs a locally finite stratification in the tame case as opposed to a finite stratification.

5.6 Applications in tame variational analysis

A standard application of Theorem 5.1 is to take first the closure of the graph of S, and then deduce generic continuity for the obtained set-valued map. While this operation is convenient, this new set-valued map no longer reflects the same local properties. For example, for a set $C \subset \mathbb{R}^n$, consider the Hadamard normal cone mapping $\hat{N}_C : \partial C \rightrightarrows \mathbb{R}^n$ and the limiting normal cone mapping $N_C : \partial C \rightrightarrows \mathbb{R}^n$, where $\mathrm{cl}(\mathrm{gph}\,(\hat{N}_C)) = \mathrm{gph}\,(N_C)$. The Hadamard normal cone $\hat{N}_C(\bar{z})$ for $\bar{z} \in \partial C$ depends on how C behaves at $\bar{z}$, whereas the normal cone $N_C(\bar{z})$ offers instead an

aggregate information from points around $\bar{z}$. The following result is comparable with [80, Proposition 6.49], and is a straightforward application of Theorem 5.19.

Corollary 5.20. *[Generic regularity]Given closed semi-algebraic sets C and D with $D \subset C$, the set-valued map $\hat{N}_C : C \rightrightarrows \mathbb{R}^n$ is continuous on $D \backslash D'$, where D' is semi-algebraic and $\dim(D') < \dim(D)$. When $D = \partial C$, we deduce that $\hat{N}_C(z) = N_C(z)$ for all z in $(\partial C) \backslash C'$, with $\dim(C') < \dim(\partial C)$.*

An analogous statement of the above corollary can be made for (nonsmooth) tangent cones $\hat{T}_C$ and T_C as well.

Remark. From the definition of subdifferential of a lower semicontinuous function [80, Definition 8.3], we can deduce that the regular (Frchet) subdifferentials are continuous outside a set of smaller dimension. This result is comparable with [80, Exercise 8.54]. Therefore nonsmoothness in tame functions and sets is structured.

Let us finally make another connection to functions whose graph is a finite union of polyhedra, hereafter referred to as *piecewise polyhedral functions*. Robinson [78] proved that a piecewise polyhedral function is calm (outer-Lipschitz) everywhere [80, Example 9.57], and a uniform Lipschitz constant suffices over the whole domain of the function (although this latter is not explicitly stated therein). A straightforward application of Theorem 5.19 yields that piecewise polyhedral functions are set-valued continuous outside a set of small dimension. We now show that a uniform Lipschitz constant for strict continuity applies.

Proposition 5.21. *[Uniformity of graphical modulus] Let $S : X \rightrightarrows \mathbb{R}^m$ be a piecewise polyhedral set-valued map, where $X \subset \mathbb{R}^n$. Then S is strictly continuous outside a set X', with $\dim(x') < \dim(x)$. Moreover, there exists some $\bar{\kappa} > 0$ such that if S is strictly continuous at $\bar{x}$, then the graphical modulus $\mathrm{lip}_X S(\bar{x} \mid \bar{y})$ is a*

nonnegative real number smaller than $\bar{\kappa}$.

Proof. The first part of the proposition of strict continuity is a direct consequence of Theorem 5.19 since S is clearly semi-algebraic. We proceed to prove the statement on the graphical modulus. We first consider the case where the graph of S is a convex polyhedron. The graph of S can be written as a finitely constrained set $\operatorname{gph}(S) = \{z \in \mathbb{R}^{n+m} \mid Az = b, Cz \leq d\}$ for some matrices A, C with finitely many rows. The projection of $\operatorname{gph}(S)$ onto $\mathbb{R}^n$ is the domain of S, which we can again write as $\operatorname{dom}(S) = X = \{x \in \mathbb{R}^n \mid A'x = b', C'x \leq d'\}$. Let $\mathcal{L}$ be the lineality space of $\operatorname{dom}(S)$, which is the set of vectors orthogonal to the rows of A'. We seek to find a constant $\bar{\kappa} > 0$ such that if x lies in the relative interior (in the sense of convex analysis) of X and $y \in S(x)$, then $\operatorname{lip}_X S(x \mid y) \leq \bar{\kappa}$. By Proposition 5.9, we have

$$\bar{\kappa} = \sup_{(x,y) \in \text{r-int}(x)} \left\{ \frac{|a|}{|b|} \mid (a,b) \in N_{\operatorname{gph}(S)}(x,y) \cap (\mathcal{L} \times \mathbb{R}^m) \right\},$$

where "r-int" stands for the relative interior. The above value is finite because of two reasons. Firstly, if $(a, \mathbf{0}) \in N_{\operatorname{gph}(S)}(x,y) \cap (\mathcal{L} \times \mathbb{R}^m)$, then by the convexity of $\operatorname{gph}(S)$, x lies on the relative boundary of X. Secondly, the "sup" in the formula is attained and can be replaced by "max". This is because the normal cones of $\operatorname{gph}(S)$ at $z = (x,y)$ can be deduced from the rows of C in which $Cz \leq d$ is actually an equation, of which there are only finitely many possibilities. In the case where S is a union of finitely many polyhedra, we consider the set-valued maps denoted by each of these polyhedra. The maximum of the Lipschitz constants for strict continuity on each polyhedral domain gives us the required $\bar{\kappa}$. □

CHAPTER 6

LEVEL SET METHODS FOR FINDING CRITICAL POINTS OF MOUNTAIN PASS TYPE

Outline: Section 6.1 illustrates a local algorithm to find saddle points of mountain pass type, while Sections 6.2, 6.3 and 6.4 are devoted to the statement, proof of convergence, and additional observations of a fast local algorithm to find nondegenerate critical points of Morse index 1 in $\mathbb{R}^n$.

Sections 6.5 discusses the relationship between mountain passes, saddle points, and critical points in the sense of metric critcal point theory and nonsmooth analysis, and does not depend on material in Sections 6.2, 6.3 and 6.4.

Finally, Sections 6.6 and 6.7 illustrates the fast local algorithm in Section 6.2. Section 6.8 discusses optimality conditions for the subproblem in the algorithm in Section 6.1.

Notation: As we will encounter situations where we want to find the square of the jth coordinate of the ith iterate of x, we write $x_i^2(j)$ in the proof of Theorem 6.12. In other parts, it will be clear from context whether the i in x_i is used as an iteration counter or as a reference to the ith coordinate. Let $\mathbb{B}^d(\mathbf{0}, r)$ be the ball with center $\mathbf{0}$ and radius r in $\mathbb{R}^d$, and $\mathring{\mathbb{B}}^d(\mathbf{0}, r)$ be the corresponding open ball.

6.1 A level set algorithm

We present a level set algorithm to find saddle points. Assume $f : X \to \mathbb{R}$, where (X, d) is a metric space.

Algorithm 6.1. *(Level set algorithm) A local bisection method for approximating*

a mountain pass from x_0 to y_0 for $f|_U$, where both x_0 and y_0 lie in some open path connected set U.

1. Start with an upper bound u and a lower bound l for the objective of the mountain pass problem and $i = 0$.

2. Solve the optimization problem

$$\begin{aligned} \min \quad & d(x, y) \\ \text{s.t.} \quad & x \in S_1, y \in S_2 \end{aligned} \tag{6.1.1}$$

   where S_1 is the component of the level set $(\text{lev}_{\leq \frac{1}{2}(l+u)} f) \cap U$ that contains x_i and S_2 is the component that contains y_i.

3. If S_1 and S_2 are the same component, then $\frac{1}{2}(l+u)$ is an upper bound, otherwise it is a lower bound. Update the upper and lower bounds accordingly. In the case where the lower bound is changed, increase i by 1, and let x_i and y_i be the minimizers of (6.1.1). For future discussions, let l_i corresponding value of l to x_i and y_i. Repeat step 2 until x_i and y_i are sufficiently close.

4. If an actual approximate mountain pass is desired, take a path $p_i : [0, 1] \to U \cap (\text{lev}_{\leq u} f)$ connecting the points

$$x_0, x_1, \ldots, x_{i-2}, x_{i-1}, x_i, y_i, y_{i-1}, y_{i-2}, \ldots, y_1, y_0.$$

Step (3) is illustrated in Figure 6.1.

To start the algorithm, an upper bound u can be taken to be the maximum of any path from x_0 to y_0, while a lower bound can be the maximum of $f(x_0)$ and $f(y_0)$. In fact, in step (3), we may update the upper bound u to be the maximum along the line segment joining x_i and y_i if it is a better upper bound.

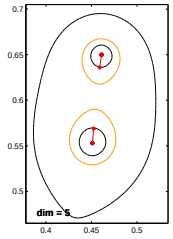
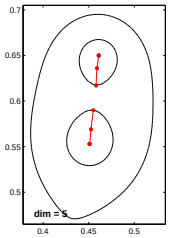
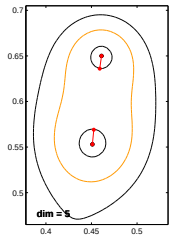
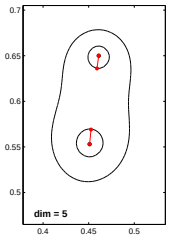
Case	Before	After
$\{x \mid f(x) \leq \frac{u+l}{2}\}$ 2 components		
1 component		

Figure 6.1: Illustration of Algorithm 6.1.

In practice, one need not solve subproblem (6.1.1) in step 2 too accurately, as it might be more profitable to move on to step 3. While theory demands the global optimizers for subproblem (6.1.1), an implementation of Algorithm 6.1 can only find local optimizers, which is not sufficient for the global mountain pass problem, but can be successful for the purpose of finding saddle points. Notice that the saddle point property is local. If x_i and y_i converge to a common limit, then it is clear from the definitions that the common limit is a saddle point.

Another issue with subproblem (6.1.1) in step 2 is that minimizers may not exist. For example, the sets S_1 and S_2 may not be compact. We now discuss how convergence to a critical point in Algorithm 6.1 can fail in the finite dimensional case.

The Palais-Smale condition is important in nonlinear analysis, and is often a necessary condition in the smooth and nonsmooth mountain pass theorems and

other critical point existence theorems. We refer to [69, 75, 77, 82, 90] for more details. We recall its definition.

Definition 6.2. Let X be a Banach space and $f : X \to \mathbb{R}$ be a $\mathcal{C}^1$ functional. We say that a sequence $\{x_i\}_{i=1}^{\infty} \subset X$ is a *Palais-Smale sequence* if $\{f(x_i)\}_{i=1}^{\infty}$ is bounded and $f'(x_i) \to \mathbf{0}$, and f satisfies the *Palais-Smale condition* if any Palais-Smale sequence admits a convergent subsequence.

For nonsmooth f, the condition $f'(x_i) \to \mathbf{0}$ is changed to $\inf_{x_i^* \in \partial f(x_i)} |x_i^*| \to 0$.

In the absence of the Palais-Smale condition, Algorithm 6.1 may fail to converge because the sequence $\{(x_i, y_i)\}_{i=1}^{\infty}$ need not have a limit point of the form $(\bar{z}, \bar{z})$, or the sequence $\{(x_i, y_i)\}_{i=1}^{\infty}$ need not even exist. The examples below document the possibilities.

Example 6.3. (a) Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = e^{-x} - y^2$. Here, the distance between the two components of the level sets is zero for all $\text{lev}_{\leq c} f$, where $c < 0$, and x_i and y_i do not exist. The sequence $\{(i, 0)\}_{i=1}^{\infty}$ is a Palais-Smale sequence but does not converge.

(b) For $f(x, y) = e^{-2x} - y^2 e^{-x}$, x_i and y_i exist, but both $\{x_i\}_{i=1}^{\infty}$ and $\{y_i\}_{i=1}^{\infty}$ do not have finite limits. Again, $\{(i, 0)\}_{i=1}^{\infty}$ is a Palais-Smale sequence that does not converge.

It is possible that $\{x_i\}_{i=1}^{\infty}$ and $\{y_i\}_{i=1}^{\infty}$ have limit points but not a common limit point. To see this, consider the example $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x & \text{if } x \leq -1 \\ -1 & \text{if } -1 \leq x \leq 1 \\ -x & \text{if } x \geq 1. \end{cases}$$

The set $\text{lev}_{\leq -1} f$ is path-connected, but the set $\text{cl}(\text{lev}_{<-1} f)$ is not path-connected. Any point in the set $(\text{lev}_{\leq -1} f) \backslash \text{cl}(\text{lev}_{<-1} f) = (-1, 1)$ is a local minimum, and hence a critical point.

6.2 A locally superlinearly convergent algorithm

In this section, we propose a locally superlinearly convergent algorithm for the mountain pass problem for smooth critical points in $\mathbb{R}^n$. For this section, we take $X = \mathbb{R}^n$. Like Algorithm 6.1 earlier, we keep track of only two points in the space $\mathbb{R}^n$ instead of a path. Our fast locally convergent algorithm does not require one to calculate the Hessian. Furthermore, we maintain upper and lower bounds that converge superlinearly to the critical value. The numerical performance of this method will be illustrated in Section 6.7.

In Algorithm 6.4 below, we can assume that the endpoints x_0 and y_0 satisfy $f(x_0) = f(y_0)$. Otherwise, if $f(x_0) < f(y_0)$ say, replace x_0 by the point x_0' closest to x_0 on the line segment $[x_0, y_0]$ such that $f(x_0') = f(y_0)$.

Algorithm 6.4. *(Fast local level set algorithm) Find saddle point between points x_0 and y_0 for $f : \mathbb{R}^n \to \mathbb{R}$. Assume that the objective of the mountain pass problem between x_0 and y_0 is greater than $f(x_0)$, and $f(x_0) = f(y_0)$. Let U be a convex set containing x_0 and y_0.*

1. Given points x_i and y_i, find z_i as follows:

   (a) Replace x_i and y_i by $\tilde{x}_i$ and $\tilde{y}_i$, where $\tilde{x}_i$ and $\tilde{y}_i$ are minimizers of the

problem

$$\begin{aligned} \min_{x,y} \quad & |x-y| \\ \text{s.t.} \quad & x \text{ lies in the same component as } x_i \text{ in } (\mathrm{lev}_{\leq f(x_i)} f) \cap U \\ & y \text{ lies in the same component as } y_i \text{ in } (\mathrm{lev}_{\leq f(x_i)} f) \cap U \end{aligned}$$

(b) Find a minimizer of f on $L_i \cap U$, say z_i. Here L_i is the affine space orthogonal to $x_i - y_i$ passing through $\frac{1}{2}(x_i + y_i)$.

2. Find the point furthest away from x_i on the line segment $[x_i, z_i]$, which we call x_{i+1}, such that $f(x) \leq f(z_i)$ for all x in the line segment $[x_i, x_{i+1}]$. Do the same to find y_{i+1}.
3. Increase i, repeat steps 1 and 2 until $|x_i - y_i|$ is small, or if the value $M_i - f(z_i)$, where $M_i := \max_{x \in [x_i, y_i]} f(x)$, is small.
4. If an actual path is desired, take a path $p_i : [0,1] \to X$ lying in $\mathrm{lev}_{\leq M_i} f$ connecting the points

$$x_0, x_1, \ldots, x_{i-2}, x_{i-1}, x_i, y_i, y_{i-1}, y_{i-2}, \ldots, y_1, y_0.$$

As we will see in Propositions 6.7 and 6.16, a unique minimizing pair $(\tilde{x}_i, \tilde{y}_i)$ in step 1(a) exists under added conditions. Furthermore, Proposition 6.3.5 implies that a unique minimizer exists of f on $L_i \cap U$ exists under added conditions in step 1(b).

To motivate step 1(b), consider any path from x_i to y_i in U that lies wholly in U. Such a path has to pass through some point of $L_i \cap U$, so the maximum value of f on the path is at least the minimum of f on $L_i \cap U$.

Step 1(a) is analogous to step 2 of Algorithm 6.1. Algorithm 6.4 can be seen as an improvement Algorithm 6.1: The bisection algorithm in Algorithm 6.1 gives us

a reliable way of finding the critical point, and step 1(b) in Algorithm 6.4 reduces the distance between the components of the level sets as fast as possible.

In practice, step 1(a) is difficult, and is performed only when the algorithm runs into difficulties. In fact, this step was not performed in our numerical experiments in Section 6.7. However, we can construct simple functions for which the affine space L_i does not separate the two components containing x_i and y_i in $(\mathrm{lev}_{\leq f(x_i)} f) \cap U$ in step 1(b) if step 1(a) were not performed.

In the minimum distance problem in step 1(a), notice that if f is $\mathcal{C}^1$ and the gradients of f at a pair of points are nonzero and do not point in opposite directions, then in principle we can perturb the points along paths that decrease the distance between them while not increasing their function values. Of course, a good approximation of a minimizing pair may be hard to compute in practice: existing path-based algorithms for finding mountain passes face analogous computational challenges. One may employ the heuristic in Remark 6.19 for this problem.

In step 2, continuity of f and p tells us that $f(x_{i+1}) = f(z_i)$. We shall see in Theorem 6.12 that under added conditions, $\{f(x_i)\}_i$ is an increasing sequence that converges to the critical value $f(\bar{x})$. Furthermore, Propositions 6.9 and 6.15 state that under added conditions, $\{M_i\}_i$ are upper bounds on $f(\bar{x})$ that converge R-superlinearly to $f(\bar{x})$.

6.3 Superlinear convergence of the local algorithm

When $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic whose Hessian has one negative eigenvalue and $n-1$ positive eigenvalues, Algorithm 6.4 converges to the critical point in one step.

One might expect that if f is $\mathcal{C}^2$, then Algorithm 6.4 converges quickly. In this section, we will prove Theorem 6.12 on the superlinear convergence of Algorithm 6.4.

Recall that the *Morse index* of a critical point is the maximum dimension of a subspace on which the Hessian is negative definite, and a critical point is *nondegenerate* if its Hessian is invertible, and degenerate otherwise. In the smooth finite dimensional case, the Morse index equals the number of negative eigenvalues of the Hessian. If a function $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in a neighborhood of a nondegenerate critical point $\bar{x}$ of Morse index 1, we can readily make the following assumptions.

Assumption 6.5. *Assume that $\bar{x} = \mathbf{0}$ and $f(\mathbf{0}) = 0$, and the Hessian $H = H(\mathbf{0})$ is a diagonal matrix with entries $a_1, a_2, \ldots, a_{n-1}, a_n$ in decreasing order, of which a_n is negative and a_{n-1} is the smallest positive eigenvalue.*

Another assumption that we will use quite often in this section and the next is on the local approximation of f near $\mathbf{0}$.

Assumption 6.6. *For $\delta \in (0, \min\{a_{n-1}, -a_n\})$, assume $\theta > 0$ is small enough so that*

$$\left| f(x) - \sum_{j=1}^{n} a_j x^2(j) \right| \leq \delta \left| x \right|^2 \text{ for all } x \in \mathbb{B}(\mathbf{0}, \theta).$$

This particular choice of θ gives a region $\mathbb{B}(\mathbf{0}, \theta)$ where Figure 6.2 is valid. We shall use $\mathring{\mathbb{B}}$ to denote the open ball.

Here is our first result on step 1(a) of Algorithm 6.4.

Proposition 6.7. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$, and $\bar{x}$ is a nondegenerate critical point of Morse index 1 such that $f(\bar{x}) = c$. If $\theta > 0$ is sufficiently small, then for any $\epsilon > 0$ (depending on θ) sufficiently small,*

1. $(\text{lev}_{\leq c-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$ *has exactly two path connected components, and*

2. *There is a pair* $(\tilde{x}, \tilde{y})$*, where* $\tilde{x}$ *and* $\tilde{y}$ *lie in distinct components of* $(\text{lev}_{\leq c-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$*, such that* $|\tilde{x} - \tilde{y}|$ *is the distance between the two components in* $(\text{lev}_{\leq c-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$.

Proof. Suppose that Assumption 6.5 holds. Choose some $\delta \in (0, \min\{a_{n-1}, -a_n\})$ and a corresponding $\theta > 0$ such that Assumption 6.6 holds. A simple bound on $f(x)$ on $\mathbb{B}(\mathbf{0}, \theta)$ is therefore:

$$\sum_{j=1}^{n} (a_j - \delta) x^2(j) \leq f(x) \leq \sum_{j=1}^{n} (a_j + \delta) x^2(j). \tag{6.3.1}$$

So if ϵ is small enough, the level set $S := \text{lev}_{\leq -\epsilon} f$ satisfies

$$S_+ \cap \mathbb{B}(\mathbf{0}, \theta) \subset S \cap \mathbb{B}(\mathbf{0}, \theta) \subset S_- \cap \mathbb{B}(\mathbf{0}, \theta),$$

where

$$\begin{aligned} S_+ &:= \left\{ x \mid \sum_{j=1}^{n} (a_j + \delta) x^2(j) \leq -\epsilon \right\}, \\ S_- &:= \left\{ x \mid \sum_{j=1}^{n} (a_j - \delta) x^2(j) \leq -\epsilon \right\}, \end{aligned}$$

and $S_+ \cap \mathbb{B}(\mathbf{0}, \theta)$ is nonempty. Figure 6.2 shows a two-dimensional cross section of the sets S_+ and S_- through the critical point $\mathbf{0}$ and the closest points between components in S_+ and S_-.

Step 1: Calculate variables in Figure 6.2.

The two points in distinct components of S_+ closest to each other are the points $\left(\mathbf{0}, \pm\sqrt{\frac{\epsilon}{-a_n - \delta}}\right)$, and one easily calculates the values of b and c (which are the distances between $\mathbf{0}$ and S_-, and that of $\mathbf{0}$ and S_+ respectively) in the diagram to be $\sqrt{\frac{\epsilon}{-a_n + \delta}}$ and $\sqrt{\frac{\epsilon}{-a_n - \delta}}$. Thus the distance between the two components of

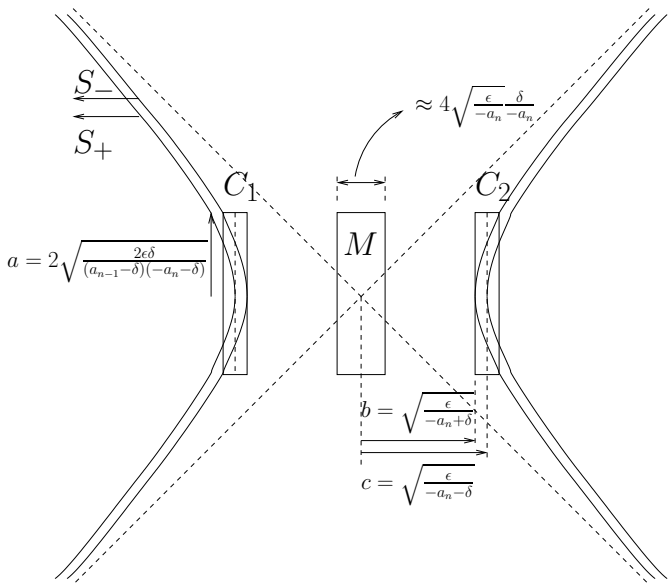


Figure 6.2: Local structure of saddle point.

S is at most $2\sqrt{\frac{\epsilon}{-a_n-\delta}}$. The points in S that minimize the distance between the components must lie in two cylinders C_1 and C_2 defined by

$$\begin{aligned} C_1 &:= \mathbb{B}^{n-1}(\mathbf{0}, a) \times [b - 2c, -b] \subset \mathbb{R}^{n-1} \times \mathbb{R}, \\ C_2 &:= \mathbb{B}^{n-1}(\mathbf{0}, a) \times [b, 2c - b] \subset \mathbb{R}^{n-1} \times \mathbb{R}, \end{aligned} \tag{6.3.2}$$

for some $a > 0$. In other words, C_1 and C_2 are cylinders with spherical base of radius a such that

$$(S_- \backslash S_+) \cap \left(\mathbb{R}^{n-1} \times [b - 2c, 2c - b]\right) \cap \mathbb{B}(\mathbf{0}, \theta) \subset C_1 \cup C_2.$$

They are represented as the left and right rectangles in Figure 6.2.

We now find a value of a. We can let $x(n) = 2c - b$, and we need

$$\begin{aligned} \sum_{j=1}^{n-1}(a_j - \delta)x^2(j) + (a_n - \delta)x^2(n) &\leq -\epsilon \\ \Rightarrow \sum_{j=1}^{n-1}(a_j - \delta)x^2(j) + (a_n - \delta)\left(2\sqrt{\frac{\epsilon}{-a_n - \delta}} - \sqrt{\frac{\epsilon}{-a_n + \delta}}\right)^2 &\leq -\epsilon. \end{aligned}$$

Continuing the arithmetic gives

$$\begin{aligned}
&\sum_{j=1}^{n-1}(a_j-\delta)x^2(j)\\
\leq\ & \epsilon\left(-1-(a_n-\delta)\left(\frac{4}{-a_n-\delta}+\frac{1}{-a_n+\delta}-\frac{4}{\sqrt{-a_n-\delta}\sqrt{-a_n+\delta}}\right)\right)\\
\leq\ & \epsilon\left(-1-(a_n-\delta)\left(\frac{4}{-a_n-\delta}+\frac{1}{-a_n+\delta}-\frac{4}{-a_n+\delta}\right)\right)\\
=\ & \frac{8\epsilon\delta}{-a_n-\delta}.
\end{aligned}$$

The radius is maximized when $x(1)=x(2)=\cdots=x(n-2)=0$ and $x(n-1)=2\sqrt{\frac{2\epsilon\delta}{(a_{n-1}-\delta)(-a_n-\delta)}}$, which gives our value of a.

Step 2: $(\mathrm{lev}_{\leq-\epsilon}f)\cap\mathring{\mathbb{B}}(\mathbf{0},\theta)$ has exactly two components if ϵ is small enough.

Observe that $(\mathrm{lev}_{\leq-\epsilon}f)\cap\mathbb{B}(\mathbf{0},\theta)$ does not intersect the subspace $L':=\{x\mid x(n)=0\}$, since $f(x)\geq 0$ for all $x\in L'\cap\mathbb{B}(\mathbf{0},\theta)$. We proceed to show that

$$U_<:=\{x\mid x(n)<0\}\cap\mathring{\mathbb{B}}(\mathbf{0},\theta)$$

contains exactly one path connected component if ϵ is small enough. A similar statement for $U_>$ defined in a similar way will allow us to conclude that $(\mathrm{lev}_{\leq-\epsilon}f)\cap\mathring{\mathbb{B}}(\mathbf{0},\theta)$ has exactly two components.

Consider two points v_1,v_2 in $(\mathrm{lev}_{\leq-\epsilon}f)\cap U_<$. We want to find a path connecting v_1 and v_2 and contained in $(\mathrm{lev}_{\leq-\epsilon}f)\cap U_<$. We may assume that $v_1(n)\leq v_2(n)<0$. By the continuity of the Hessian, assume that θ is small enough so that for all $x\in\mathbb{B}(\mathbf{0},\theta)$, the top left principal submatrix of $H(x)$ corresponding to the first $n-1$ elements is positive definite. Consider the subspace $L'(\alpha):=\{x\mid x(n)=\alpha\}$.

The positive definiteness of the submatrix of $H(x)$ on $\mathbb{B}(\mathbf{0},\theta)$ tells us that f is strictly convex on $\mathbb{B}(\mathbf{0},\theta) \cap L'(\alpha)$.

If $v_1(n) = v_2(n)$, then the line segment connecting v_1 and v_2 lies in $(\text{lev}_{\leq -\epsilon} f) \cap L'(v_1(n)) \cap \mathring{\mathbb{B}}(\mathbf{0},\theta)$ by the convexity of f on $L'(v_1(n)) \cap \mathring{\mathbb{B}}(\mathbf{0},\theta)$. Otherwise, assume that $v_1(n) < v_2(n)$.

Here is a lemma that we will need for the proof.

Lemma 6.8. *Suppose Assumption 6.5 holds. We can reduce $\theta > 0$ and $\delta > 0$ if necessary so that Assumption 6.6 is satisfied, and the nth component of $\nabla f(x)$ is positive for all $x \in (\text{lev}_{\leq 0} f) \cap \mathbb{B}(\mathbf{0},\theta) \cap \{x \mid x(n) < 0\}$.*

Proof. We first define $\tilde{S}_-$ by

$$\tilde{S}_- := \{x \mid (a_{n-1} - \delta) \sum_{j=1}^{n-1} x^2(j) + (a_n - \delta) x^2(n) \leq 0\}.$$

It is clear that $(a_{n-1} - \delta) \sum_{j=1}^{n-1} x^2(j) + (a_n - \delta) x^2(n) \leq f(x)$ for all $x \in \mathbb{B}(\mathbf{0},\theta)$, so $(\text{lev}_{\leq 0} f) \cap \mathbb{B}(\mathbf{0},\theta) \subset \tilde{S}_- \cap \mathbb{B}(\mathbf{0},\theta)$.

We now use the expansion $\nabla f(x) = H(\mathbf{0})x + o(|x|)$, and prove that the nth component of $\nabla f(x)$ is negative for all $x \in \tilde{S}_- \cap \mathbb{B}(\mathbf{0},\theta) \cap \{x \mid x(n) < 0\}$. We can reduce θ so that $|\nabla f(x) - H(\mathbf{0})x| < \delta |x|$ for all $x \in \mathbb{B}(\mathbf{0},\theta)$. Note that if $x \in \tilde{S}_-$, then

$$\begin{aligned} (a_{n-1} - \delta) \sum_{j=1}^{n-1} x^2(j) + (a_n - \delta) x^2(n) &\leq 0 \\ \Rightarrow (a_{n-1} - \delta) |x|^2 + (a_n - a_{n-1}) x^2(n) &\leq 0 \\ \Rightarrow |x| &\leq \sqrt{\frac{a_{n-1} - a_n}{a_{n-1} - \delta}} \, (-x(n)) \, . \end{aligned}$$

The nth component of $\nabla f(x)$ is bounded from below by

$$a_n x(n) - \delta |x| \leq a_n x(n) + \delta \sqrt{\frac{a_{n-1} - a_n}{a_{n-1} - \delta}} x(n).$$

Provided that δ is small enough, the term above is positive since $x(n) < 0$. $\square$

We now return to show that there is a path connecting v_1 and v_2. Note that $S_+ \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta) \cap \{x \mid x(n) < 0\}$ is a convex set. (To see this, note that $S_+ \cap \{x \mid x(n) < 0\}$ can be rotated so that it is the epigraph of a convex function.) Since $S_+ \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta) \subset (\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$, the open line segment connecting the points $(\mathbf{0}, -\theta), (\mathbf{0}, -c) \in \mathbb{R}^{n-1} \times \mathbb{R}$ lies in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$. If $-\theta < v_1(n) < v_2(n) \leq -c$, the piecewise linear path connecting v_2 to $(\mathbf{0}, v_2(n))$ to $(\mathbf{0}, v_1(n))$ to v_1 lies in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$.

In the case when $v_2(n) > -c$, we see that v_2 must lie in C_1. Lemma 6.8 tells us that the line segment joining v_2 and $v_2 + (\mathbf{0}, -c - v_2(n))$ lies in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$. This allows us to find a path connecting v_2 to v_1.

Step 3: $\tilde{x}$ and $\tilde{y}$ lie in $\mathring{\mathbb{B}}(\mathbf{0}, \theta)$.

The points $\tilde{x}$ and $\tilde{y}$ must lie in C_1 and C_2 respectively, and both C_1 and C_2 lie in $\mathring{\mathbb{B}}(\mathbf{0}, \theta)$ if ϵ is small enough. Therefore, we can minimize over the compact sets $(\text{lev}_{\leq -\epsilon} f) \cap C_1$ and $(\text{lev}_{\leq -\epsilon} f) \cap C_2$, which tells us that a minimizing pair $(\tilde{x}, \tilde{y})$ exist. $\square$

In fact, under the assumptions of Proposition 6.7, $\tilde{x}$ and $\tilde{y}$ are unique, but all we need in the proof of Proposition 6.9 below is that $\tilde{x}$ and $\tilde{y}$ lie in the sets C_1 and C_2 defined by (6.3.2) respectively ans represented as rectangles in Figure 6.2. We defer the proof of uniqueness to Proposition 6.16.

Our next result is on a bound for possible locations of z_i in step 1(b).

Proposition 6.9. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$, and $\bar{x}$ is a nondegenerate critical point of Morse index 1 such that $f(\bar{x}) = c$. If θ is small enough, then for*

all small $\epsilon > 0$ (depending on θ),

(1) Two closest points of the two components of $(\text{lev}_{\leq c-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$, say $\tilde{x}$ and $\tilde{y}$, exist,

(2) For any such points $\tilde{x}$ and $\tilde{y}$, f is strictly convex on $L \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$, where L is the orthogonal bisector of $\tilde{x}$ and $\tilde{y}$, and

(3) f has a unique minimizer on $L \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$. Furthermore, $\min_{L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)} f \leq f(\bar{x}) \leq \max_{[\tilde{x}, \tilde{y}]} f$.

Proof. Suppose that Assumption 6.5 holds, and choose $\delta \in (0, \min\{a_{n-1}, -a_n\})$. Suppose that $\theta > 0$ is small enough such that Assumption 6.6 holds. Throughout this proof, we assume all vectors accented with a hat '$\wedge$' are of Euclidean length 1. It is clear that $f(\tilde{x}) = f(\tilde{y}) = -\epsilon$. Point (1) of the result comes from Proposition 6.7. We first prove the following lemma.

Lemma 6.10. *Suppose Assumptions 6.5 and 6.6 hold. If $\theta > 0$ is small enough, then for all small $\epsilon > 0$ (depending on θ), two closest points of the two components of $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$, say $\tilde{x}$ and $\tilde{y}$, exist. Let L be the perpendicular bisector of $\tilde{x}$ and $\tilde{y}$. Then*

$$(\text{lev}_{\leq 0} f) \cap L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta) \subset \mathbb{B}^{n-1}\left(\mathbf{0}, \alpha\sqrt{\frac{(-a_n + \delta)}{(a_{n-1} - \delta)}}\right) \times (-\alpha, \alpha)\,,$$
$$\text{where } \alpha = \delta\sqrt{\frac{\epsilon}{-a_n}}\left(\frac{8}{a_{n-1}} + \frac{2}{-a_n}\right) + o(\delta).$$

Proof. By Proposition 6.7, the points $\tilde{x}$ and $\tilde{y}$ must exist. We proceed to prove the rest of Lemma 6.10.

Step 1: Calculate remaining values in Figure 6.2.

We calculated the values of a, b and c in step 2 of the proof of Proposition 6.7, and we proceed to calculate the rest of the variables in Figure 6.2. The middle rectangle in Figure 6.2 represents the possible locations of midpoints of points in C_1 and C_2, and is a cylinder as well. We call this set M. The radius of this cylinder is the same as that of C_1 and C_2, and the width of this cylinder is $4(c-b)$, which gives an $o(\delta)$ approximation

$$\begin{aligned} 4(c-b) &= 4\left(\sqrt{\frac{\epsilon}{-a_n-\delta}}-\sqrt{\frac{\epsilon}{-a_n+\delta}}\right) \\ &= 4\sqrt{\frac{-a_n\epsilon}{(-a_n-\delta)(-a_n+\delta)}}\left(\sqrt{1+\frac{\delta}{-a_n}}-\sqrt{1-\frac{\delta}{-a_n}}\right) \\ &= 4\sqrt{\frac{\epsilon}{-a_n}}\left(\left(1+\frac{\delta}{-2a_n}\right)-\left(1-\frac{\delta}{-2a_n}\right)\right)+o(\delta) \\ &= 4\sqrt{\frac{\epsilon}{-a_n}}\frac{\delta}{-a_n}+o(\delta). \end{aligned}$$

These calculations suffice for the calculations in step 2 of this proof.

Step 2: Set up optimization problem for bound on $(\mathrm{lev}_{\leq 0}f)\cap L\cap\mathring{\mathbb{B}}(\mathbf{0},\theta)$.

From the values of a and b calculated previously, we deduce that a vector c_2-c_1, with $c_i\in C_i$, can be scaled so that it is of the form $(\gamma\frac{a}{b}\hat{\mathbf{v}}_1,1)$, where $\hat{\mathbf{v}}_1\in\mathbb{R}^{n-1}$ is of norm 1 and $0\leq\gamma\leq 1$. (i.e., the norm corresponding to the first $n-1$ coordinates is at most $\frac{a}{b}$.) These are possible normals for L, the perpendicular bisector of $\tilde{x}$ and $\tilde{y}$. The formula for $\frac{a}{b}$ is

$$\begin{aligned} \frac{a}{b} &= 2\sqrt{\frac{2\epsilon\delta}{(a_{n-1}-\delta)(-a_n-\delta)}}\div\sqrt{\frac{\epsilon}{-a_n+\delta}} \\ &= 2\sqrt{\frac{2\delta(-a_n+\delta)}{(a_{n-1}-\delta)(-a_n-\delta)}}. \end{aligned}$$

So we can represent a normal of the affine space L as

$$\left(2\gamma_1\sqrt{\frac{2\delta(-a_n+\delta)}{(a_{n-1}-\delta)(-a_n-\delta)}}\hat{\mathbf{v}}_1,1\right)\text{ for some }0\leq\gamma_1\leq 1. \tag{6.3.3}$$

We now proceed to bound the minimum of f on all possible perpendicular bisectors of c_1 and c_2 within $\mathring{\mathbb{B}}(\mathbf{0}, \theta)$, where $c_1 \in C_1$ and $c_2 \in C_2$. We find the largest value of α such that

- there is a point of the form $(\mathbf{v}_2, \alpha)$ lying in $\tilde{S}_-$, where

$$\tilde{S}_- := \{x \mid (a_{n-1} - \delta) \sum_{j=1}^{n-1} x^2(j) + (a_n - \delta)x^2(n) \leq 0\} \subset \mathbb{R}^{n-1} \times \mathbb{R}.$$

- $(\mathbf{v}_2, \alpha) \in \tilde{L}$ for some affine space $\tilde{L}$ passing through a point $p \in M$ and having a normal vector of the form in Formula (6.3.3).

The set $\tilde{S}_-$ is the same as that defined in the proof of Lemma 6.8. Note that $\tilde{S}_- \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta) \supset (\text{lev}_{\leq 0} f) \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$, and this largest value of α is an upper bound on the absolute value of the nth coordinate of elements in $(\text{lev}_{\leq 0} f) \cap L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$.

Step 3: Solving for α.

For a point $(\mathbf{v}_2, \alpha) \in \tilde{S}_-$, where $\mathbf{v}_2 = (x(1), x(2), \ldots, x(n-1)) \in \mathbb{R}^{n-1}$, we have

$$\begin{aligned}
(a_{n-1} - \delta) \sum_{j=1}^{n-1} x^2(j) + (a_n - \delta)\alpha^2 &\leq 0. \\
\Rightarrow |\mathbf{v}_2|^2 &= \sum_{j=1}^{n-1} x^2(j) \\
&\leq \frac{(-a_n + \delta)}{(a_{n-1} - \delta)} \alpha^2. \\
\Rightarrow |\mathbf{v}_2| &\leq \sqrt{\frac{(-a_n + \delta)}{(a_{n-1} - \delta)}} \alpha.
\end{aligned}$$

Therefore, we can write $(\mathbf{v}_2, \alpha)$ as

$$\left(\gamma_2 \sqrt{\frac{(-a_n + \delta)}{(a_{n-1} - \delta)}} \alpha \hat{\mathbf{v}}_2, \alpha \right), \tag{6.3.4}$$

where $\hat{\mathbf{v}}_2 \in \mathbb{R}^{n-1}$ is a vector of unit norm, and $0 \leq \gamma_2 \leq 1$. We can assume that p has coordinates

$$\left(2\gamma_3\sqrt{\frac{2\epsilon\delta}{(a_{n-1}-\delta)(-a_n-\delta)}}\hat{\mathbf{v}}_3, 2\gamma_4\sqrt{\frac{\epsilon}{-a_n}}\frac{\delta}{-a_n} + o(\delta)\right),$$

where $\hat{\mathbf{v}}_3 \in \mathbb{R}^{n-1}$ is some vector of unit norm, and $0 \leq \gamma_3, \gamma_4 \leq 1$. Note that the nth component is half the width of M. Hence a possible tangent on $\tilde{L}$ is

$$\left(\gamma_1\sqrt{\frac{(-a_n+\delta)}{(a_{n-1}-\delta)}}\alpha\hat{\mathbf{v}}_2, \alpha\right) - \left(2\gamma_3\sqrt{\frac{2\epsilon\delta}{(a_{n-1}-\delta)(-a_n-\delta)}}\hat{\mathbf{v}}_3, 2\gamma_4\sqrt{\frac{\epsilon}{-a_n}}\frac{\delta}{-a_n} + o(\delta)\right).$$

To simplify notation, note that we only require an $O(\delta)$ approximation of α, we can take the terms like $-a_n + \delta$ and $-a_n - \delta$ to be $-a_n + O(\delta)$ and so on. The dot product of the above vector and the normal of the affine space L calculated in Formula (6.3.3) must be zero, which after some simplification gives:

$$\begin{aligned}\left(\left(\gamma_2\sqrt{\frac{-a_n}{a_{n-1}}} + O(\delta)\right)\alpha\hat{\mathbf{v}}_2 - \left(2\gamma_3\sqrt{\frac{2\epsilon\delta}{a_{n-1}(-a_n)}} + O(\delta^{3/2})\right)\hat{\mathbf{v}}_3 \right.\\ \left., \alpha - \left(2\gamma_4\sqrt{\frac{\epsilon}{-a_n}}\frac{\delta}{-a_n} + o(\delta)\right)\right) \cdot \left(\left(2\gamma_1\sqrt{\frac{2\delta}{a_{n-1}}} + O(\delta^{3/2})\right)\hat{\mathbf{v}}_1, 1\right) &= 0.\end{aligned}$$

At this point, we remind the reader that the $O(\delta^k)$ terms mean that there exists some $K > 0$ such that if δ were small enough, we can find terms t_1 to t_3 such that $|t_i| < K\delta^k$ and the formula above is satisfied by t_i in place of the $O(\delta^k)$ terms. Further arithmetic gives

$$\begin{aligned}&4\gamma_1\gamma_3\sqrt{\frac{2\delta}{a_{n-1}}}\sqrt{\frac{2\epsilon\delta}{a_{n-1}(-a_n)}}(\hat{\mathbf{v}}_3 \cdot \hat{\mathbf{v}}_1) + 2\gamma_4\sqrt{\frac{\epsilon}{-a_n}}\frac{\delta}{-a_n} + o(\delta)\\ &= \alpha\left(1 + 2\gamma_1\gamma_2\sqrt{\frac{2\delta}{a_{n-1}}}\sqrt{\frac{-a_n}{a_{n-1}}}(\hat{\mathbf{v}}_2 \cdot \hat{\mathbf{v}}_1) + o(\delta^{3/2})\right)\\ &= \alpha(1 + O(\sqrt{\delta}))\end{aligned}$$

To find an upper bound for α, it is clear that we should take $\gamma_1 = \gamma_3 = \gamma_4 = 1$ and $\hat{\mathbf{v}}_3 \cdot \hat{\mathbf{v}}_1 = 1$. The $O(\sqrt{\delta})$ term is superfluous, and this simplifies to give

$$\alpha \leq \delta\sqrt{\frac{\epsilon}{-a_n}}\left(\frac{8}{a_{n-1}} + \frac{2}{-a_n}\right) + o(\delta). \tag{6.3.5}$$

We could find the minimum possible value of α by these same series of steps and show that the absolute value would be bounded above by the same bound. This ends the proof of Lemma 6.10. $\square$

It is clear that the minimum value of f on $L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$ is at most 0, since L intersects the axis corresponding to the nth coordinate and f is nonpositive there. Therefore the set $(\text{lev}_{\leq 0} f) \cap L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$ is nonempty, and f has a local minimizer on $L \cap \mathring{\mathbb{B}}(\mathbf{0}, \theta)$.

We now state and prove our second lemma that will conclude the proof of Proposition 6.9.

Lemma 6.11. *Let L be the perpendicular bisector of $\tilde{x}$ and $\tilde{y}$ as defined in point (1) of Proposition 6.9 with $\bar{x} = \mathbf{0}$. If δ and θ are small enough satisfying Assumptions 6.5 and 6.6, then $f\,|_{L \cap \mathbb{B}(\mathbf{0}, \theta)}$ is strictly convex.*

Proof. The lineality space of L, written as $\text{lin}(L)$, is the space of vectors orthogonal to $\tilde{x} - \tilde{y}$. We can infer from Formula (6.3.3) that $\tilde{x} - \tilde{y}$ is a scalar multiple of a vector of the form $(w, 1)$, where $w \in \mathbb{R}^{n-1}$ satisfies $|w| \to \mathbf{0}$ as $\delta \to 0$. We consider a vector $v \in \text{lin}(L)$ orthogonal to $(w, 1)$ that can be scaled so that $v = (\tilde{w}, 1)$, where $(w, 1) \cdot (\tilde{w}, 1) = 0$, which gives $w \cdot \tilde{w} = -1$. The Cauchy Schwarz inequality gives us

$$\begin{aligned} |\tilde{w}|\,|w| &\geq |\tilde{w} \cdot w| \\ &= 1 \\ \Rightarrow |\tilde{w}| &\geq |w|^{-1}. \end{aligned}$$

So

$$\begin{aligned}\frac{v^\top H(p)v}{v^\top v} &= \frac{v^\top H(\mathbf{0})v}{v^\top v} + \frac{v^\top (H(p)-H(\mathbf{0}))v}{v^\top v} \\ &= \frac{\sum_{j=1}^{n-1} a_j v^2(j) + a_n}{\sum_{j=1}^{n-1} v^2(j) + 1} + \frac{v^\top (H(p)-H(\mathbf{0}))v}{v^\top v} \\ &\geq \underbrace{\frac{a_{n-1}\sum_{j=1}^{n-1} v^2(j) + a_n}{\sum_{j=1}^{n-1} v^2(j) + 1}}_{(1)} + \underbrace{\frac{v^\top (H(p)-H(\mathbf{0}))v}{v^\top v}}_{(2)}.\end{aligned}$$

Since $\sum_{j=1}^{n-1} v^2(j) = |\tilde{w}|^2 \to \infty$ as $|w| \to 0$, the limit of term (1) is a_{n-1}, so there is an open set $\mathbb{B}(\mathbf{0},\theta)$ containing $\mathbf{0}$ such that $\frac{v^\top H(p)v}{v^\top v} > \frac{1}{2}a_{n-1}$ for all $v \in \mathrm{lin}(L) \cap \{x \mid x(n) = 1\}$ and $p \in \mathbb{B}(\mathbf{0},\theta)$. By the continuity of the Hessian, we may reduce θ if necessary so that $\|H(p) - H(\mathbf{0})\| < \frac{1}{2}a_{n-1}$ for all $p \in \mathbb{B}(\mathbf{0},\theta)$. Thus $\frac{v^\top H(p)v}{v^\top v} > 0$ for all $p \in \mathbb{B}(\mathbf{0},\theta)$ and $v \in \mathrm{lin}(L) \cap \{x \mid x(n) = 1\}$ if δ is small enough.

The vectors of the form $v = (\tilde{w}, 0)$ do not present additional difficulties as the corresponding term (1) is at least a_{n-1}. This proves that the Hessian $H(p)$ restricted to $\mathrm{lin}(L)$ is positive definite, and hence the strict convexity of f on $L \cap \mathring{\mathbb{B}}(\mathbf{0},\theta)$. □

Since f has a local minimizer in $L \cap \mathring{\mathbb{B}}(\mathbf{0},\theta)$ and is strictly convex there, we have (2) and the first part of part (3). The inequality $f(\bar{x}) \leq \max_{[\tilde{x},\tilde{y}]} f$ follows easily from the fact that the line segment $[\tilde{x},\tilde{y}]$ intersects the set $\{x \mid x(n) = 0\}$, on which f is nonnegative. □

Here is our theorem on the convergence of Algorithm 6.4.

Theorem 6.12. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in a neighborhood of a nondegenerate critical point $\bar{x}$ of Morse index 1. If $\theta > 0$ is sufficiently small and x_0 and y_0 are chosen such that*

(a) x_0 *and* y_0 *lie in the two different components of* $(\mathrm{lev}_{\leq f(x_0)} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$,

(b) $f(x_0) = f(y_0) < f(\bar{x})$,

then Algorithm 6.4 with $U = \mathring{\mathbb{B}}(\bar{x}, \theta)$ *generates a sequence of iterates* $\{\tilde{x}_i\}_i$ *and* $\{\tilde{y}_i\}_i$ *lying in* $\mathring{\mathbb{B}}(\bar{x}, \theta)$ *such that the function values* $\{f(\tilde{x}_i)\}_i$ *and* $\{f(\tilde{y}_i)\}_i$ *converge to* $f(\bar{x})$ *superlinearly, and the iterates* $\{\tilde{x}_i\}_i$ *and* $\{\tilde{y}_i\}_i$ *converge to* $\bar{x}$ *superlinearly.*

Proof. As usual, suppose Assumption 6.5 holds, and δ and θ are chosen so that Assumption 6.6 holds.

Step 1: Linear convergence of $f(\tilde{x}_i)$ to critical value $f(\bar{x})$.

Let $\epsilon = f(\tilde{x}_i)$. The next iterate x_{i+1} satisfies $f(x_{i+1}) = f(z_i)$, and is bounded from below by

$$f(x_{i+1}) \geq (a_n - \delta)\alpha^2 = -\epsilon\delta^2 \left(\frac{8}{a_{n-1}} + \frac{2}{-a_n} \right)^2 + o(\delta^2),$$

where α is the value calculated in Lemma 6.10. The ratio between the previous function value and the next function value is at most

$$\rho(\delta) := \delta^2 \left(\frac{8}{a_{n-1}} + \frac{2}{-a_n} \right)^2 + o(\delta^2).$$

This ratio goes to 0 as $\delta \searrow 0$, so we can choose some δ small enough so that $\rho < \frac{1}{2}$. We can choose θ corresponding to the value of δ satisfying Assumption 6.6. This shows that the convergence to 0 of the function values $f(\tilde{x}_{i+1}) = f(x_{i+1})$ in Algorithm 6.4 is linear provided x_0 and y_0 lie in $\mathbb{B}(\mathbf{0}, \theta)$ and ϵ is small enough by Proposition 6.7. We can reduce θ if necessary so that $f(x) \geq -\epsilon$ for all $x \in \mathbb{B}(\mathbf{0}, \theta)$, so the condition on ϵ does not present difficulties.

Step 2: Superlinear convergence of $f(\tilde{x}_i)$ to critical value $f(\bar{x})$.

Choose a sequence $\{\delta_k\}_k$ so that $\delta_k \searrow 0$ monotonically. Corresponding to δ_k, we can choose θ_k satisfying Assumption 6.6. Since $\{\tilde{x}_i\}_i$ and $\{\tilde{y}_i\}_i$ converge to $\mathbf{0}$, for any $k \in \mathbb{Z}_+$, we can find some $i^* \in \mathbb{Z}_+$ so that the cylinders C_1 and C_2 constructed in Figure 6.2 corresponding to $\epsilon_i = -f(\tilde{x}_i)$ and $\delta = \delta_1$ lie wholly in $\mathbb{B}(\mathbf{0}, \theta_k)$ for all $i > i^*$. As remarked in step 3 of the proof of Proposition 6.7, $\tilde{x}_i$ and $\tilde{y}_i$ must lie inside C_1 and C_2, so we can take $\delta = \delta_k$ for the ratio ρ. This means that $\frac{|f(\tilde{x}_{i+1})|}{|f(\tilde{x}_i)|} \leq \rho(\delta_k)$ for all $i > i^*$. As $\rho(\delta) \searrow 0$ as $\delta \searrow 0$, this means that we have superlinear convergence of the $f(\tilde{x}_i)$ to the critical value $f(\bar{x})$.

Step 3: Superlinear convergence of $\tilde{x}_i$ to the critical point $\bar{x}$.

We now proceed to prove that the distance between the critical point $\mathbf{0}$ and the iterates decrease superlinearly by calculating the value $\frac{|\tilde{x}_{i+1}|}{|\tilde{x}_i|}$, or alternatively $\frac{|\tilde{x}_{i+1}|^2}{|\tilde{x}_i|^2}$. The value $|\tilde{x}_i|$ satisfies $|\tilde{x}_i|^2 \geq b^2 = \frac{\epsilon}{-a_n+\delta}$. To find an upper bound for $|\tilde{x}_{i+1}|^2$, it is instructive to look at an upper bound for $|\tilde{x}_i|^2$ first. As can be deduced from Figure 6.2, an upper bound for $|\tilde{x}_i|^2$ is the square of the distance between $\mathbf{0}$ and the furthest point in C_1, which is

$$\begin{aligned}(2c-b)^2 + a^2 &= (c + (c-b))^2 + a^2 \\ &= \frac{\epsilon}{-a_n - \delta} + 8\frac{\epsilon\delta}{(-a_n)^2} + \frac{8\epsilon\delta}{(a_{n-1}-\delta)(-a_n-\delta)} + o(\delta).\end{aligned}$$

This means that an upper bound for $|\tilde{x}_{i+1}|^2$ is

$$\delta^2 \left(\frac{8}{a_{n-1}} + \frac{2}{-a_n}\right)^2 \left(\frac{\epsilon}{-a_n-\delta} + \frac{8\epsilon\delta}{-a_n}\left(\frac{1}{-a_n} + \frac{1}{(a_{n-1}-\delta)}\right)\right) + o(\delta^2).$$

From this point, one easily sees that as $i \to \infty$, $\delta \to 0$, and $\frac{|\tilde{x}_{i+1}|^2}{|\tilde{x}_i|^2} \to 0$. This gives the superlinear convergence of the distance between the critical point and the iterates $\tilde{x}_i$ that we seek. $\square$

6.4 Further properties of the local algorithm

In this section, we take note of some interesting properties of Algorithm 6.4. First, we show that it is easy to find x_{i+1} and y_{i+1} in step 2 of Algorithm 6.4.

Proposition 6.13. *Suppose the conditions in Theorem 6.12 hold. Consider the sequence of iterates $\{x_i\}_i$ and $\{y_i\}_i$ generated by Algorithm 6.4. If i is large enough, then either $x_{i+1} = z_i$ or $y_{i+1} = z_i$ in step 2 of Algorithm 6.4.*

Proof. Let $\tilde{p} : [0, 1] \to \mathbb{R}^n$ denote the piecewise linear path connecting x_i to z_i to y_i. It suffices to prove that along $\tilde{p}$, the function f increases to a maximum, and then decreases. Suppose Assumptions 6.5 and 6.6 hold. The cylinders C_1 and C_2 in Figure 6.2 are loci for x_i and y_i. We assume that x_i lies in C_2 in Figure 6.2. The calculations in (6.3.4) in Lemma 6.10 tell us that z_i can be written as

$$\left(\sqrt{\frac{(-a_n + \delta)}{(a_{n-1} - \delta)}} \alpha \lambda_1 \hat{\mathbf{v}}_2, \lambda_2 \alpha \right) \in \mathbb{R}^{n-1} \times \mathbb{R},$$

where $0 < \lambda_1 < \lambda_2 \leq 1$, $|\hat{\mathbf{v}}_2| = 1$ and $\alpha = \delta \sqrt{\frac{\epsilon}{-a_n}} \left(\frac{8}{a_{n-1}} + \frac{2}{-a_n} \right) + o(\delta)$ by (6.3.5). Therefore, $x_i - z_i$ can be written as

$$\left(\mathbf{v}_1, \sqrt{\frac{\epsilon}{-a_n + \delta}} + o(\sqrt{\delta\epsilon}) \right),$$

where $\mathbf{v}_1 \in \mathbb{R}^{n-1}$ satisfies

$$\begin{aligned} |\mathbf{v}_1| &\leq \sqrt{\frac{(-a_n + \delta)}{(a_{n-1} - \delta)}} \alpha + a \\ &= O(\sqrt{\epsilon\delta}), \end{aligned}$$

and $a = \sqrt{\frac{2\epsilon\delta}{(a_{n-1} - \delta)(-a_n - \delta)}}$ is as calculated in the proof of Proposition 6.7. This means that the unit vector with direction $x_i - z_i$ converges to the n-th elementary vector as $\delta \searrow 0$. By appealing to Hessians as is done in the proof of Lemma 6.11,

we see that the function f is strictly concave in the line segment $[x_i, z_i]$ if i is large enough. Similarly, f is strictly concave in the line segment $[y_i, z_i]$ if i is large enough.

Next, we prove that the function f has only one local maximizer in $\tilde{p}([0,1])$. In the case where $\nabla f(z_i) = \mathbf{0}$, the concavity of f on the line segments $[x_i, z_i]$ and $[y_i, z_i]$ tells us that z_i is the a unique maximizer on $\tilde{p}([0,1])$. We now look at the case where $\nabla f(z_i) \neq \mathbf{0}$. Since z_i is the minimizer on a subspace with normal $x_i - y_i$, $\nabla f(z_i)$ is a (possibly negative) multiple of $x_i - y_i$. This means that $\nabla f(z_i) \cdot (x_i - z_i)$ has a different sign than $\nabla f(z_i) \cdot (y_i - z_i)$. In other words, the map $t \mapsto f(\tilde{p}\,(t))$ increases then decreases. This concludes the proof of the proposition. □

Remark 6.14. Note that in Algorithm 6.4, all we need in step 1 is a good lower bound of the critical value. We can exploit convexity as proved in Lemma 6.11 and use cutting plane methods to attain a lower bound for f on $L_i \cap \mathbb{B}(\bar{x}, \theta)$.

Recall from Proposition 6.9 that M_i is a sequence of upper bounds of the critical value $f(\bar{x})$. While it is not even clear that M_i is monotonically decreasing, we can prove the following convergence result on M_i. Recall that a sequence in $\mathbb{R}$ converges *R-superlinearly* to zero if its absolute value is bounded by a superlinearly convergent sequence.

Proposition 6.15. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in a neighborhood of a nondegenerate critical point $\bar{x}$ of Morse index 1, the neighborhood U of $\bar{x}$ and the points x_0 and y_0 are chosen satisfying the conditions in the statement of Theorem 6.12. Then in Algorithm 6.4, $M_i := \max_{[x_i, y_i]} f$ converges R-superlinearly to the critical value.*

Proof. Suppose Assumption 6.5 holds. An upper bound of the critical value of the

saddle point is obtained by finding the maximum along the line segment joining two points in C_1 and C_2, which is bounded from above by

$$(a_1 + \delta)a^2 \;\; = \;\; (a_1 + \delta)\frac{8\epsilon\delta}{(a_{n-1} - \delta)(-a_n - \delta)}.$$

A more detailed analysis by using cylinders with ellipsoidal base instead of circular base tell us that the maximum is bounded above by $\frac{8\epsilon\delta}{(-a_n-\delta)}$ instead. If $\delta > 0$ is small enough, this value is much smaller than $-f(x_i) = \epsilon$. As $i \to \infty$, the estimates $-f(x_i)$ converge superlinearly to 0 by Theorem 6.12, giving us what we need.

□

Step 1(a) is important in the analysis of Algorithm 6.4. As explained earlier in Section 6.2, it may be difficult to implement this step. Algorithm 6.4 may run fine without ever performing step 1(a) (see the example in Section 6.7), but it may need to be performed occasionally in a practical implementation. The following result tells us that under the assumptions we have made so far, this problem is locally a strictly convex problem with a unique solution.

Proposition 6.16. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$ in a neighborhood of a nondegenerate critical point $\bar{x}$ of Morse index 1 with critical value $f(\bar{x}) = c$. Then if $\epsilon > 0$ is small enough, there is a convex neighborhood U_ϵ of $\bar{x}$ such that $(\text{lev}_{\leq c-\epsilon} f) \cap U_\epsilon$ is a union of two disjoint convex sets.*

Consequently, providing θ is sufficiently small, the pair of nearest points guaranteed by Proposition 6.7(2) are unique.

Proof. Suppose Assumptions 6.5 and 6.6 hold. In addition, we further assume that

$$|\nabla f(x) - H(x)| < \delta\,|x| \;\text{ for all } x \in \mathring{\mathbb{B}}(\mathbf{0}, \theta).$$

We can choose U_ϵ to be the interior of $\mathrm{conv}(C_1 \cup C_2)$, where C_1 and C_2 are the cylinders in Figure 6.2 and defined in the proof of Proposition 6.7, but in view of Theorem 6.18, we shall prove that U_ϵ can be chosen to be the bigger set $\mathrm{conv}(\tilde{C}_1 \cup \tilde{C}_2)$, where $\tilde{C}_1$ and $\tilde{C}_2$ are cylinders defined by

$$\begin{aligned} \tilde{C}_1 &:= \mathbb{B}^{n-1}(\mathbf{0}, \rho) \times [-\beta, -b] \subset \mathbb{R}^{n-1} \times \mathbb{R}, \\ \tilde{C}_2 &:= \mathbb{B}^{n-1}(\mathbf{0}, \rho) \times [b, \beta] \subset \mathbb{R}^{n-1} \times \mathbb{R}, \end{aligned}$$

where β, ρ are constants to be determined. We choose β such that

$$\mathbb{B}^{n-1}(\mathbf{0}, a) \times \{\beta\} \subset \mathrm{int}(S_+).$$

In particular, β satisfies

$$\begin{aligned} a^2(a_1 + \delta) + \beta^2(a_n + \delta) &< -\epsilon \\ \Rightarrow \beta^2 &> \frac{1}{-a_n - \delta}\left(\epsilon + a^2(a_1 + \delta)\right) \\ &= \frac{\epsilon}{-a_n - \delta}\left(1 + \frac{8\delta(a_1 + \delta)}{(a_{n-1} - \delta)(-a_n - \delta)}\right) \end{aligned}$$

We choose β to be any value satisfying the above inequality.

Next, we choose ρ to be the smallest value such that $S_- \cap (\mathbb{R}^{n-1} \times [-\beta, \beta]) \cap \mathbb{B}(\mathbf{0}, \theta) \subset \tilde{C}_1 \cup \tilde{C}_2$. This calculation is similar to the calculation of a, which gives

$$\begin{aligned} (a_{n-1} - \delta)\rho^2 + (a_n - \delta)\beta^2 &= -\epsilon \\ \Rightarrow \rho &= \sqrt{\frac{-\epsilon - (a_n - \delta)\beta^2}{a_{n-1} - \delta}}. \end{aligned}$$

We shall not expand the terms, but remark that β and ρ are of $O(\sqrt{\epsilon})$.

The proof of Proposition 6.7 tells us that $\mathrm{conv}(\tilde{C}_1 \cup \tilde{C}_2) \cap \mathrm{lev}_{\leq -\epsilon} f$ is a union of the two nonempty sets $\tilde{C}_1 \cap \mathrm{lev}_{\leq -\epsilon} f$ and $\tilde{C}_2 \cap \mathrm{lev}_{\leq -\epsilon} f$. It remains to show that these two sets are strictly convex.

Any point $x \in \tilde{C}_1$ can be written as

$$x = (\mathbf{x}', x_n),$$

where $\mathbf{x}' \in \mathbb{R}^{n-1}$ is of norm at most ρ, and $-\beta \leq x_n \leq -b$, where β is as calculated above and $b = \sqrt{\frac{\epsilon}{-a_n+\delta}}$ as in Figure 6.2. This implies that

$$Hx = (\mathbf{x}'', a_n x_n),$$

where $\mathbf{x}''$ is of norm at most $a_1 |\mathbf{x}'|$. It is clear that as $\delta \downarrow 0$, the unit vector in the direction of Hx converges to $(\mathbf{0}, 1)$. This implies that for any $\kappa_1 > 0$, there exists some $\delta > 0$ such that $\text{unit}(\nabla f(x)) \cdot (\mathbf{0}, 1) \geq 1 - \kappa_1$ for all $x \in C_1$. (Note that $\tilde{C}_1$ depends on δ.) Here, unit : $\mathbb{R}^n \backslash \{\mathbf{0}\} \to \mathbb{R}^n$ is the mapping of a nonzero vector to the unit vector pointing in the same direction.

Let z_1 and z_2 be points in $\tilde{C}_1 \cap (\text{lev}_{\leq -\epsilon} f)$. Suppose that $z_1(n) < z_2(n)$, and let $\mathbf{v} = (\mathbf{v}_1, v_2) \in \mathbb{R}^{n-1} \times \mathbb{R}$ be a unit vector in the same direction as $z_2 - z_1$. We further assume, by reducing θ and δ as necessary, that $\|H(x) - H(\mathbf{0})\| < \kappa_2$ for all $x \in \tilde{C}_1 \cap (\text{lev}_{\leq -\epsilon} f)$. Suppose κ_1 and κ_2 are small enough so that $\sqrt{2\kappa_1} < \sqrt{\frac{a_{n-1}-\kappa_2}{a_{n-1}-a_n}}$.

Note that $v_2 \geq 0$. Either one of these two cases on v_2 must hold. We prove that in both cases, he open line segment (z_1, z_2) lies in the interior of $(\text{lev}_{\leq -\epsilon} f) \cap \tilde{C}_1$.

Case 1: $v_2 > \sqrt{2\kappa_1}$**.**

In this case, for all $x \in \tilde{C}_1$, we have

$$\begin{aligned}
\mathbf{v} \cdot (\text{unit}(\nabla f(x))) &= \mathbf{v} \cdot (\mathbf{0}, 1) + \mathbf{v} \cdot (\text{unit}(\nabla f(x)) - (\mathbf{0}, 1)) \\
&\geq v_2 - |\mathbf{v}| \, |\text{unit}(\nabla f(x)) - (\mathbf{0}, 1)| \\
&= v_2 - |\text{unit}(\nabla f(x)) - (\mathbf{0}, 1)| \\
&= v_2 - \sqrt{|\text{unit}(\nabla f(x))|^2 + |(\mathbf{0}, 1)|^2 - 2\text{unit}(\nabla f(x)) \cdot (\mathbf{0}, 1)} \\
&> v_2 - \sqrt{2 - 2(1 - \kappa_1)} \\
&= v_2 - \sqrt{2\kappa_1} \\
&> 0.
\end{aligned}$$

This means that along the line segment $[z_1, z_2]$, the function f is strictly monotone. Therefore, if $x_1, x_2 \in (\text{lev}_{\leq -\epsilon} f) \cap \tilde{C}_1$, the open line segment (z_1, z_2) lies in the interior of $(\text{lev}_{\leq -\epsilon} f) \cap \tilde{C}_1$.

Case 2: $v_2 < \sqrt{\frac{a_{n-1} - \kappa_2}{a_{n-1} - a_n}}$**.**

Let $H^u(\mathbf{0})$ denote the diagonal matrix of size $(n-1) \times (n-1)$ with elements $a_1, \ldots, a_{n-1}$. We have

$$\begin{aligned}
\mathbf{v}^\top H(x) \mathbf{v} &= \mathbf{v}^\top H(\mathbf{0}) \mathbf{v} + \mathbf{v}^\top (H(x) - H(\mathbf{0})) \mathbf{v} \\
&> \mathbf{v}_1^\top H^u(\mathbf{0}) \mathbf{v}_1 + a_n v_2^2 - |\mathbf{v}|^2 \, \|H(x) - H(\mathbf{0})\| \\
&\geq a_{n-1} |\mathbf{v}_2|^2 + a_n v_2^2 - \|H(x) - H(\mathbf{0})\| \\
&> a_{n-1}(1 - v_2^2) + a_n v_2^2 - \kappa_2 \\
&= a_{n-1} + v_2^2 (a_n - a_{n-1}) - \kappa_2 \\
&> a_{n-1} + (\kappa_2 - a_{n-1}) - \kappa_2 \\
&\geq 0
\end{aligned}$$

This means that the function f is strictly convex along the line segment $[z_1, z_2]$, so if $x_1, x_2 \in (\text{lev}_{\leq -\epsilon} f) \cap \tilde{C}_1$, the open line segment (z_1, z_2) lies in the interior of

$(\text{lev}_{\leq -\epsilon} f) \cap \tilde{C}_1$, concluding the proof of the first part of this result.

To prove the next statement on the uniqueness of the pair of closest points, suppose that $(\tilde{x}', \tilde{y}')$ and $(\tilde{x}'', \tilde{y}'')$ are distinct pairs whose distance give the distance between the components of $(\text{lev}_{\leq -\epsilon} f) \cap \mathbb{B}(\mathbf{0}, \theta)$, where $\mathbb{B}(\mathbf{0}, \theta)$ is as stated in Proposition 6.7. If ϵ is small enough, then $\text{conv}(\tilde{C}_1 \cup \tilde{C}_2)$ lies in $\mathring{\mathbb{B}}(\mathbf{0}, \theta)$. Then by the strict convexity of the components of $(\text{lev}_{\leq -\epsilon} f) \cap \text{conv}(\tilde{C}_1 \cup \tilde{C}_2)$, the pair $(\frac{1}{2}(\tilde{x}' + \tilde{x}''), \frac{1}{2}(\tilde{y}' + \tilde{y}''))$ lie in the same components, and the distance between this pair of points must be the same as that for the pairs $(\tilde{x}', \tilde{y}')$ and $(\tilde{x}'', \tilde{y}'')$. The closest points in the components of $[\frac{1}{2}(\tilde{x}' + \tilde{x}''), \frac{1}{2}(\tilde{y}' + \tilde{y}'')] \cap \text{lev}_{\leq -\epsilon} f$ give a smaller distance between the components of $(\text{lev}_{\leq -\epsilon} f) \cap \mathbb{B}(\mathbf{0}, \theta)$, which contradicts the optimality of the pairs $(\tilde{x}', \tilde{y}')$ and $(\tilde{x}'', \tilde{y}'')$. □

Note that in the case of $\epsilon = 0$, there may be no neighborhood U_0 of $\bar{x}$ such that $U_0 \cap (\text{lev}_{\leq c} f)$ is a union of two convex sets intersecting only at the critical point. We also note that U_ϵ depends on ϵ in our result above. The following example explains these restrictions.

Example 6.17. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = (x_2 - x_1^2)(x_1 - x_2^2)$. The shaded area in Figure 6.3 is a sketch of $\text{lev}_{\leq 0} f$.

We now explain that the neighborhood U_ϵ defined in Proposition 6.16 must depend on ϵ for this example. For any open U containing $\mathbf{0}$, we can always find two points p and q in a component of $(\text{lev}_{<0} f) \cap U$ such that the line segment $[p, q]$ does not lie in $\text{lev}_{<0} f$. This implies that the component of $(\text{lev}_{\leq -\epsilon} f) \cap U$ is not convex if $0 < \epsilon \leq -\max(f(p), f(q))$. ⋄

We now take a second look at the problem of minimizing the distance between two components in step 1(a) of Algorithm 6.4. We need to solve the following

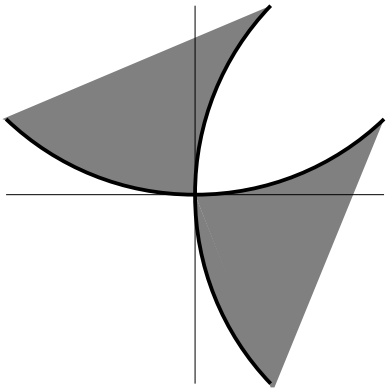


Figure 6.3: $\text{lev}_{\leq 0} f$ for $f(x) = (x_2 - x_1^2)(x_1 - x_2^2)$

problem for $\epsilon > 0$:

$$\begin{aligned} \min_{x,y} \quad & |x - y| \\ \text{s.t.} \quad & x \text{ in same component as } a \text{ in } (\text{lev}_{\leq f(\bar{x})-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta) \\ & y \text{ in same component as } b \text{ in } (\text{lev}_{\leq f(\bar{x})-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta). \end{aligned} \tag{6.4.1}$$

If $(\tilde{x}, \tilde{y})$ is a pair of local optimizers, then $\tilde{y}$ is the closest point to the component of $(\text{lev}_{\leq f(\bar{x})-\epsilon} f) \cap U$ containing $\tilde{x}$ and vice versa. This gives us the following optimality conditions:

$$\begin{aligned} \nabla f(\tilde{x}) &= \kappa_1(\tilde{y} - \tilde{x}), \\ \nabla f(\tilde{y}) &= \kappa_2(\tilde{x} - \tilde{y}), \\ f(\tilde{x}) &= f(\bar{x}) - \epsilon \\ f(\tilde{y}) &= f(\bar{x}) - \epsilon \\ \text{for some } \kappa_1, \kappa_2 &\geq 0. \end{aligned} \tag{6.4.2}$$

From Proposition 6.16, we see that given any $\theta > 0$ sufficiently small, provided that the conditions in Proposition 6.7 hold, the global minimizing pair of (6.4.1) is unique. Even though convexity is absent, the following theorem shows that the global minimizing pair is, under added conditions, the only pair satisfying the optimality conditions (6.4.2), showing that there are no other local minimizers of (6.4.1).

Theorem 6.18. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is $\mathcal{C}^2$, and $\bar{x}$ is a nondegenerate critical point of Morse index 1 such that $f(\bar{x}) = c$. If $\theta > 0$ is sufficiently small, then for any $\epsilon > 0$ (depending on θ) sufficiently small, the global minimizer of* (6.4.1) *is the only pair in $\mathring{\mathbb{B}}(\bar{x}, \theta) \times \mathring{\mathbb{B}}(\bar{x}, \theta)$ satisfying the optimality conditions* (6.4.2)*.*

Proof. Suppose that Assumption 6.5 holds, and δ is chosen small enough so that 6.6 holds. We also assume that θ is small enough so that $|H(x) - H(\mathbf{0})| < \frac{1}{2}\min(a_{n-1}, -a_n)$. Seeking a contradiction, suppose that $(\tilde{x}, \tilde{y})$ satisfy the optimality conditions.

We refer to Figure 6.2, and also recall the definitions of the sets $\tilde{C}_1$ and $\tilde{C}_2$ in the proof of Proposition 6.16. As proven in Proposition 6.16, the convexity properties of the two level sets in $(\text{lev}_{\leq f(\bar{x})-\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$ imply that if $\tilde{x} \in \tilde{C}_1$, $\tilde{y} \in \tilde{C}_2$ and the optimality conditions are satisfied, then the pair $(\tilde{x}, \tilde{y})$ is the global minimizing pair.

Consider the case where $\tilde{x} \notin \tilde{C}_1$. Either of the two cases hold. We note the asymmetry below in that we check whether $\tilde{y} \in C_2$ instead of whether $\tilde{y} \in \tilde{C}_2$.

Case 1: $\tilde{y} \in C_2$: In this case, if the first $n-1$ coordinates of $\tilde{x}$ are the same as that of $\tilde{y}$, then $\tilde{x}$ lies in the interior of $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$, which is a contradiction to optimality. Recall that the value of β was chosen such that $\tilde{y} + (\mathbf{0}, \tilde{x}(n) - \tilde{y}(n))$ lies in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$. By the convexity of $f|_{L'(\tilde{x}(n))}$, where $L'(\tilde{x}(n))$ is the affine space $\{x \mid x(n) = \tilde{x}(n)\}$, the line segment connecting $\tilde{x}$ and $\tilde{y} + (\mathbf{0}, \tilde{x}(n) - \tilde{y}(n))$ lies in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$. The distance between $\tilde{y}$ and points along this line segment decreases (at a linear rate) as one moves away from $\tilde{x}$, which again contradicts the assumption that $(\tilde{x}, \tilde{y})$ satisfy (6.4.2).

Case 2: $\tilde{y} \notin C_2$: By the convexity of $f|_{L'(\tilde{x}(n))}$ and $f|_{L'(\tilde{y}(n))}$, the line segments

$[\tilde{y}, \tilde{y} - (\mathbf{0}, \tilde{y}(n))]$ and $[\tilde{x}, \tilde{x} - (\mathbf{0}, \tilde{x}(n))]$ lie in $(\text{lev}_{\leq -\epsilon} f) \cap \mathring{\mathbb{B}}(\bar{x}, \theta)$. These line segments and the optimality of the pair $(\tilde{x}, \tilde{y})$ implies that the first $n-1$ components of $\tilde{x}$ and $\tilde{y}$ to be the same. This in turn implies that $\nabla f(\tilde{x})$ is a positive multiple of $(\mathbf{0}, 1)$.

Our proof ends if we show that if θ is small enough, $\nabla f(\tilde{x})$ cannot be a positive multiple of $(\mathbf{0}, 1)$. If $\tilde{x} \notin \tilde{C}_1$, then $\tilde{x}(n) < -\beta$. If $\tilde{x}$ lies on the boundary of $\text{lev}_{\leq -\epsilon} f$, then $f(\tilde{x}) = -\epsilon$, and we have

$$\begin{aligned}
f(\tilde{x}) &= -\epsilon \\
\sum_{i=1}^{n} (a_i + \delta)\tilde{x}(i)^2 &\geq -\epsilon \\
(a_1 + \delta) \sum_{i=1}^{n} \tilde{x}(i)^2 + (a_n - a_1)\tilde{x}(n)^2 &\geq -\epsilon \\
(a_1 + \delta)|\tilde{x}|^2 &\geq (a_1 - a_n)\tilde{x}(n)^2 - \epsilon \\
\frac{|\tilde{x}|^2}{\tilde{x}(n)^2} &\geq \frac{a_1 - a_n - \frac{\epsilon}{\tilde{x}(n)^2}}{a_1 + \delta} \\
&\geq 1 + \frac{-a_n - \delta - \frac{\epsilon}{\beta^2}}{a_1 + \delta}
\end{aligned}$$

Upon expansion of the term β^2 in the expression in the final line, we see that $\frac{|\tilde{x}|^2}{\tilde{x}(n)^2}$ is bounded from below by a constant independent of ϵ and greater than 1. Since f is $\mathcal{C}^2$, the set

$$\{x \mid \nabla f(x) \text{ is a multiple of } (\mathbf{0}, 1)\} \cap \mathbb{B}(\mathbf{0}, \theta)$$

is a manifold, whose tangent at the origin is the line spanned by $(\mathbf{0}, 1)$. This implies that if θ is small enough, then $\tilde{x} \notin \tilde{C}_1$ and $\tilde{x}$ lying on the boundary of $\text{lev}_{\leq -\epsilon} f$ implies that $\nabla f(\tilde{x})$ cannot be a multiple of $(\mathbf{0}, 1)$. We have the required contradiction. □

Remark 6.19. We now describe a heuristic to approximate a pair of closest points iteratively between the components of $(\text{lev}_{\leq c-\epsilon} f) \cap U$. For two points x' and y'

that approximate $\tilde{x}_i$ and $\tilde{y}_i$, we can find local minimizers of f on the affine spaces orthogonal to $x' - y'$ that pass through x' and y' respectively, say x^*, y^*, and then find the closest points in the two components of $(\mathrm{lev}_{\leq c-\epsilon} f) \cap [x^*, y^*]$, where $[x^*, y^*]$ is the line segment connecting x^* and y^*. This heuristic is particularly practical in the case of Wilkinson problem, as we illuminate in Sections 6.6 and 6.7.

6.5 Saddle points and criticality properties

We have seen that Algorithm 6.1 allows us to find saddle points of mountain type. In this section, we first prove an equivalent definition of a saddle point based on paths connecting two points. Then we prove that saddle points are critical points in the metric sense and in the nonsmooth sense.

In the following equivalent condition for saddle points, we say that a path $p : [0, 1] \to X$ *connects* a and b if $p(0) = a$ and $p(1) = b$, and it is *contained in* $U \subset X$ if $p([0, 1]) \subset U$. The *maximum value* of the path p is defined as $\max_t f \circ p(t)$.

Proposition 6.20. *Let (X, d) be a metric space. For a continuous function $f : X \to \mathbb{R}$, $\bar{x}$ is a saddle point of mountain pass type if and only if there exists an open neighborhood U and two points $a, b \in (\mathrm{lev}_{<l} f) \cap U$ such that*

(a) The maximum value of any path connecting a and b contained in U is at least $f(\bar{x})$, and

(b) for all $\epsilon > 0$, there exists $\delta, \theta \in (0, \epsilon)$ and a path p_ϵ connecting a and b contained in U such that the maximum value of p_ϵ is at most $f(\bar{x}) + \epsilon$, and $(\mathrm{lev}_{\geq f(\bar{x})-\theta} f) \cap p_\epsilon([0, 1]) \subset \mathbb{B}(\bar{x}, \delta)$.

Proof. We first prove that the conditions (a) and (b) above imply that $\bar{x}$ is a saddle point. Let A and B be the path connected components of $\text{lev}_{<f(\bar{x})}f \cap U$ containing a and b respectively. For any $\epsilon > 0$, the condition $(\text{lev}_{\geq f(\bar{x})-\theta}f) \cap p_\epsilon([0,1]) \subset \mathbb{B}(\bar{x}, \delta)$ tells us that we can find points $x_\epsilon \in A$ and $y_\epsilon \in B$ such that $d(\bar{x}, x_\epsilon) < \delta < \epsilon$ and $d(\bar{x}, y_\epsilon) < \epsilon$. For a sequence $\epsilon_i \searrow 0$, we set $x_i = x_{\epsilon_i}$ and $y_i = y_{\epsilon_i}$. This shows that $\bar{x}$ lies in both the closure of A and that of B, and hence $\bar{x}$ is a saddle point.

Next, we prove the converse. Suppose that $\bar{x}$ is a saddle point, with U being a neighborhood of $\bar{x}$, and the sets A and B are two path components of $(\text{lev}_{<f(\bar{x})}f) \cap U$ whose closures contain $\bar{x}$. For any $\epsilon > 0$, we can find some $\delta \in (0, \epsilon)$ such that $d(x, \bar{x}) < \delta$ implies $|f(x) - f(\bar{x})| < \epsilon$. There are two points $x_\epsilon \in A$ and $y_\epsilon \in B$ such that $d(x_\epsilon, \bar{x}) < \delta$ and $d(y_\epsilon, \bar{x}) < \delta$.

Let a and b be any two points in the sets A and B respectively. There is a path connecting a to x_ϵ contained in $\text{lev}_{<f(\bar{x})}f \cap U$, say p_a, and we can similarly find a path p_b connecting y_ϵ to b contained in $\text{lev}_{<f(\bar{x})}f \cap U$. The maximum values on both paths p_a and p_b are less than $f(\bar{x})$, so there is some $\theta \in (0, \epsilon)$ such that both maximum values are bounded above by $f(\bar{x}) - \theta$. Choose a path p'_ϵ to be the line segment connecting x_a and y_b contained in $\mathbb{B}(\bar{x}, \delta)$. The path p_ϵ formed by the concatenation of the paths p_a, p'_ϵ and p_b satisfies condition (b). Condition (a) is easily seen to be satisfied, and hence we are done. □

Ideally, we want to improve condition (b) in Proposition 6.20 so that $\bar{x}$ is the maximum point on some mountain pass connecting a and b. We shall see in Example 6.22 that saddle points in general need not have this property. A simple finite dimensional condition on the function f so that this happens is semi-algebraicity. A set in $\mathbb{R}^n$ is *semi-algebraic* if it is a union of finitely many sets defined by finitely many polynomial inequalities, and a function $f : \mathbb{R}^n \to \mathbb{R}$

is *semi-algebraic* if its graph $\{(x, y) \in \mathbb{R}^n \times \mathbb{R} \mid y = f(x)\}$ is a semi-algebraic set. Semi-algebraic objects remove much of the oscillatory behavior that typically does not appear in applications, and form a large class of objects that appear in applications. We will appeal to semi-algebraic geometry for only the next result, and we refer readers interested in the general theory of semi-algebraic functions (and more generally, that of o-minimal structures and tame topology, under which Proposition 6.21 also holds) to [11, 34, 33, 42].

Proposition 6.21. *In the case where $f : \mathbb{R}^n \to \mathbb{R}$ is semi-algebraic, condition (b) in Proposition 6.20 can be replaced with*

(b′) There is a path connecting a and b contained in U along which the unique maximizer is $\bar{x}$.

Proof. It is clear that (b′) is a stronger condition than (b), so we prove that if f is semi-algebraic, then (b′) holds. Suppose $\bar{x}$ is a saddle point of mountain pass type. Let U be an open neighborhood of $\bar{x}$, and sets A and B be two components of $(\text{lev}_{<f(\bar{x})} f) \cap U$ whose closures contain $\bar{x}$. Choose points $a \in A$ and $b \in B$. It is clear that A and B are semi-algebraic (see for example [33, Section 3.2]. By the curve selection lemma (see for example [33, Section 3.1]), there is a path p_a connecting a and $\bar{x}$ such that $p_a(1) = \bar{x}$, and $p_a([0, 1)) \subset A$. Similarly, we can find a path p_b connecting $\bar{x}$ and b such that $p_b(0) = \bar{x}$ and $p_b((0, 1]) \subset B$. The concatenation of p_a and p_b gives us what we need. □

In the absence of semi-algebraicity, the following example illustrates that a saddle point need not satisfy condition (b′).

Example 6.22. We define $f : \mathbb{R}^2 \to \mathbb{R}$ through Figure 6.4. There are 2 shapes in the positive quadrant the figure: a blue "comb" C wrapping around a brown

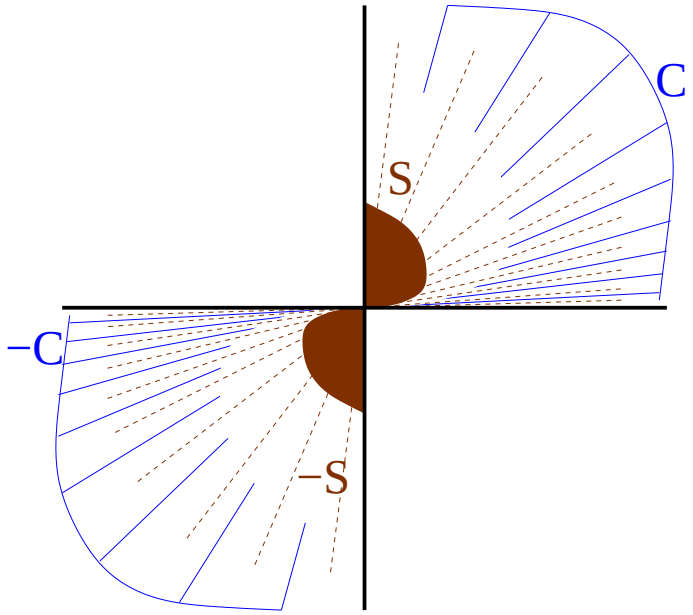


Figure 6.4: Illustration of saddle point in Example 6.22.

"sun" S. The closure of C contains the origin $\mathbf{0}$ (the intersection of the horizontal and vertical axis).

We can define a continuous $f : \mathbb{R}^2 \to \mathbb{R}$ so that f is negative on $C \cup (-C)$ and positive on $(S \cup (-S))\backslash\{\mathbf{0}\}$ and $\{(x, y) \mid xy < 0\}$, and extend f continuously to all of $\mathbb{R}^2$ using the Tietze extension theorem. It is clear that $\mathbf{0}$ is a saddle point, and the sets $A, B \subset \mathrm{lev}_{<0} f$ whose closures contain $\mathbf{0}$ can be taken to be the path connected components containing C and $(-C)$ respectively. But the origin $\mathbf{0}$ does not satisfy condition (b′).

Our next step is to establish the relation between saddle points and criticality in metric spaces. We recall the following definitions in metric critical point theory from [36, 54, 59].

Definition 6.23. Let (X, d) be a metric space. We call the point x *Morse regular*

for the function $f : X \to \mathbb{R}$ if, for some numbers $\gamma, \sigma > 0$, there is a continuous function

$$\phi : \mathbb{B}(x, \gamma) \times [0, \gamma] \to X$$

such that all points $u \in \mathbb{B}(x, \gamma)$ and $t \in [0, \gamma]$ satisfy the inequality

$$f(\phi(x, t)) \leq f(x) - \sigma t,$$

and that $\phi(\cdot, 0)$ is the identity map. The point x is *Morse critical* if it is not Morse regular.

If there is some $\kappa > 0$ and such a function ϕ that also satisfies the inequality

$$d(\phi(x, t), x) \leq \kappa t,$$

then we call x *deformationally regular*. The point x is *deformationally critical* if it is not deformationally regular.

We now relate saddle points to Morse critical and deformationally critical points.

Proposition 6.24. *For a function $f : X \to \mathbb{R}$ defined on a metric space X, $\bar{x}$ is a saddle point of mountain pass type implies that $\bar{x}$ is deformationally critical. If in addition, either $X = \mathbb{R}^n$ or condition (b′) in Proposition 6.21 holds, then $\bar{x}$ is Morse critical.*

Proof. Let U be an open neighborhood of $\bar{x}$ as defined in Definition 1.3, and let A and B be two distinct components of $(\text{lev}_{<f(\bar{x})} f) \cap U$ which contain $\bar{x}$ in their closures. The proofs of all three results by contradiction are similar. For convenience, we label the following three assumptions as follows, and prove that they all lead to the contradiction that A and B cannot be distinct path components in U.

(D) $\bar{x}$ is deformationally regular.

$(M_{\mathbb{R}^n})$ $\bar{x}$ is Morse regular, and $X = \mathbb{R}^n$.

$(M_{b'})$ $\bar{x}$ is Morse regular, and condition (b′) in Proposition 6.21 holds.

Suppose condition $(M_{\mathbb{R}^n})$ holds. Let $\gamma, \sigma > 0$ and $\phi : \mathbb{B}(\bar{x}, \gamma) \times [0, \gamma] \to X$ satisfy the properties of Morse regularity given in Definition 6.23. We can assume that γ is small enough so that $\mathbb{B}(\bar{x}, \gamma) \subset U$. By the continuity of ϕ and the compactness of $\mathbb{B}(\bar{x}, \gamma)$, there is some $\gamma' > 0$ such that $\mathbb{B}(\bar{x}, \gamma) \times [0, \gamma'] \subset \phi^{-1}(U)$.

Next, suppose condition (D) holds. Let $\gamma, \sigma, \kappa > 0$ and $\phi : \mathbb{B}(\bar{x}, \gamma) \times [0, \gamma] \to X$ satisfy the properties given in Definition 6.23 on deformation regularity. We can assume $\gamma > 0$ is small enough and choose $\gamma' > 0$ so that $\mathbb{B}(\bar{x}, \gamma + \gamma'\kappa) \subset U$. The conditions on ϕ imply that $\phi\left(\mathbb{B}(\bar{x}, \gamma) \times [0, \gamma']\right) \subset \mathbb{B}(\bar{x}, \gamma + \gamma'\kappa) \subset U$, which in turn imply that $\mathbb{B}(\bar{x}, \gamma) \times [0, \gamma'] \subset \phi^{-1}(U)$.

Here is the next argument common to both conditions (D) and $(M_{\mathbb{R}^n})$. By the characterization of saddle points in Proposition 6.20, we can find θ and δ satisfying the condition in Proposition 6.20(b) with $\theta, \delta \leq \min(\frac{1}{2}\gamma'\sigma, \gamma)$. This gives us $\mathbb{B}(\bar{x}, \delta) \subset \mathbb{B}(\bar{x}, \gamma) \subset U$ in particular. We can glean from the proof of Proposition 6.20 that we can find two points $a_\delta \in A \cap \mathbb{B}(\bar{x}, \delta)$ and $b_\delta \in B \cap \mathbb{B}(\bar{x}, \delta)$ and a path $p' : [0, 1] \to X$ connecting a_δ and b_δ contained in $\mathbb{B}(\bar{x}, \delta)$ with maximum value at most $f(\bar{x}) + \min(\frac{1}{2}\gamma'\sigma, \gamma)$. The functions values $f(a_\delta)$ and $f(b_\delta)$ satisfy $f(a_\delta), f(b_\delta) \leq f(\bar{x}) - \theta$. The condition $\mathbb{B}(\bar{x}, \gamma) \times [0, \gamma'] \subset \phi^{-1}(U)$ implies that $p'([0, 1]) \times [0, \gamma'] \subset \phi^{-1}(U)$.

If condition $(M_{b'})$ holds, then for any $\delta > 0$, we can find a path $p' : [0, 1] \to X$ connecting two points $a_\delta \in A \cap \mathbb{B}(\bar{x}, \delta)$ and $b_\delta \in B \cap \mathbb{B}(\bar{x}, \delta)$ contained in $\mathbb{B}(\bar{x}, \delta)$ with maximum value at most $f(\bar{x})$. There is also some $\theta > 0$ such that $f(a_\delta), f(b_\delta) <$

$f(\bar{x}) - \theta$. Let $\gamma, \sigma > 0$ and $\phi : \mathbb{B}(\bar{x}, \gamma) \times [0, \gamma] \to X$ be such that they satisfy the properties of Morse regularity. By the compactness of $p'([0,1])$, we can find some $\gamma' > 0$ such that $p'([0,1]) \times [0, \gamma'] \subset \phi^{-1}(U)$.

To conclude the proof for all three cases, consider the path $\bar{p} : [0,3] \to X$ defined by

$$\bar{p}(t) = \begin{cases} \phi(a_\delta, \gamma' t) & \text{for } 0 \leq t \leq 1 \\ \phi(p'(t-1), \gamma') & \text{for } 1 \leq t \leq 2 \\ \phi(b_\delta, \gamma'(3-t)) & \text{for } 2 \leq t \leq 3. \end{cases}$$

This path connects a_δ and b_δ, is contained in U and has maximum value at most $\max(f(\bar{x}) - \theta, f(\bar{x}) - \frac{1}{2}\gamma'\sigma)$, which is less than $f(\bar{x})$. This implies that A and B cannot be distinct path connected components of $(\text{lev}_{<f(\bar{x})} f) \cap U$, which establishes the contradiction in all three cases. □

We now move on to discuss how saddle points and deformationally critical points relate to nonsmooth critical points. Here is the definition of Clarke critical points.

Definition 6.25. [31, Section 2.1] Let X be a Banach space. Suppose $f : X \to \mathbb{R}$ is locally Lipschitz. The *Clarke generalized directional derivative* of f at x in the direction $v \in X$ is defined by

$$f^\circ(x; v) = \limsup_{t \searrow 0, y \to x} \frac{f(y + tv) - f(y)}{t},$$

where $y \in X$ and t is a positive scalar. The *Clarke subdifferential* of f at x, denoted by $\partial_C f(x)$, is the convex subset of the dual space X^* given by

$$\{\zeta \in X^* \mid f^\circ(x; v) \geq \langle \zeta, v \rangle \text{ for all } v \in X\}.$$

The point x is a *Clarke (nonsmooth) critical point* if $\mathbf{0} \in \partial_C f(x)$. Here, $\langle \cdot, \cdot \rangle : X^* \times X \to \mathbb{R}$ defined by $\langle \zeta, v \rangle := \zeta(v)$ is the dual relation.

For the particular case of $\mathcal{C}^1$ functions, $\partial_C f(x) = \{\nabla f(x)\}$. Therefore a critical point of a smooth function (i.e., a point x that satisfies $\nabla f(x) = \mathbf{0}$) is also a Clarke critical point. From the definitions above, it is clear that an equivalent definition of a Clarke critical point is $f^\circ(x; v) \geq 0$ for all $v \in X$. This property allows us to deduce Clarke criticality without appealing to the dual space X^*. It is well known that this definition is equivalent to Definition 2.14 in the Lipschitz case.

Clarke (nonsmooth) critical points of f are of interest in, for example, partial differential equations with discontinuous nonlinearities. Critical point existence theorems for nonsmooth functions first appeared in [28, 84]. For the problem of finding nonsmooth critical points numerically, we are only aware of [92].

The following result is well-known, and we include its proof for completeness.

Proposition 6.26. *Let X be a Banach space and $f : X \to \mathbb{R}$ be locally Lipschitz at $\bar{x}$. If $\bar{x}$ is deformationally critical, then it is Clarke critical.*

Proof. We prove the contrapositive instead. If the point $\bar{x}$ is not Clarke critical, there exists a unit vector $v \in X$ such that

$$\limsup_{t \searrow 0, y \to \bar{x}} \frac{f(y + tv) - f(y)}{t} < 0.$$

Now defining $\phi(x, t) = x - tv$ satisfies the conditions for deformation regularity. □

To conclude, Figure 6.5 summarizes the relationship between saddle points and the different types of critical points.

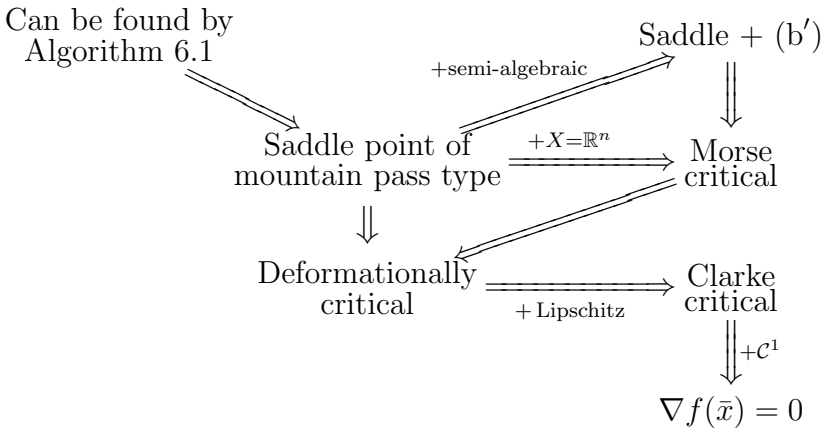


Figure 6.5: Different types of critical points

6.6 Wilkinson's problem: Background

In Section 6.7, we will apply Algorithm 6.4 to attempt to solve the *Wilkinson problem*: given a matrix A, we seek a nearest matrix A' with a multiple eigenvalue. The *Wilkinson distance* is the distance between the matrices A and A'. We should remark that solving the Wilkinson problem requires solving a global mountain pass problem, while we only attempt to find critical points of mountain pass type in this chapter. We begin by describing the Wilkinson problem in this section and our numerical results in Section 6.7.

Though not cited explicitly, as noted by [1], the Wilkinson problem can be traced back to [89, pp. 90-93]. See [25, 68] for more references, and in particular, the discussion in the beginning of [25, Section 3].

It is well-known that eigenvalues vary in a Lipschitz manner if and only if they do not coincide. In fact, eigenvalues are differentiable in the entries of the matrix when they are distinct. Hence, as discussed by Demmel [38], the Wilkinson distance is a natural condition measure for accurate eigenvalue computation. The

Wilkinson distance is also important because of its connections with the stability of eigendecompositions of matrices. To our knowledge, no fast and reliable numerical method for computing the Wilkinson distance is known.

The *ϵ-pseudospectrum* $\Lambda_\epsilon(A) \subset \mathbb{C}$ of A is defined as the set

$$\begin{aligned}\Lambda_\epsilon(A) &:= \{z \mid \exists E \text{ s.t. } \|E\| \leq \epsilon \text{ and } z \text{ is an eigenvalue of } A+E\} \\ &= \left\{z \mid \left|(A-zI)^{-1}\right|^{-1} \leq \epsilon\right\} \\ &= \{z \mid \underline{\sigma}(A-zI) \leq \epsilon\},\end{aligned}$$

where $\underline{\sigma}(A-zI)$ is the smallest singular value of $A-zI$. The function $z \mapsto (A-zI)^{-1}$ is sometimes referred to as the resolvent function, whose (Clarke) critical points are referred to as *resolvent critical points*. To simplify notation, define $\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}_+$ by

$$\begin{aligned}\underline{\sigma}_A(z) &:= \underline{\sigma}(A-zI) \\ &= \text{smallest singular value of } (A-zI).\end{aligned}$$

For more on pseudospectra, we refer the reader to [86].

It is well known that each component of the ϵ-pseudospectrum $\Lambda_\epsilon(A)$ contains at least one eigenvalue. If ϵ is small enough, $\Lambda_\epsilon(A)$ has n components, each containing an eigenvalue. Let $\bar{\epsilon}$ be the smallest ϵ for which $\Lambda_\epsilon(A)$ contains $n-1$ components or less. Alam and Bora [1] proved that $\bar{\epsilon}$ is the Wilkinson distance for A. Two components of $\Lambda_\epsilon(A)$ would coalesce when $\epsilon \uparrow \bar{\epsilon}$, and the point at which two components coalesce can be used to construct the matrix with repeated eigenvalues closest to A. Equivalently, the point of coalescence of the two components is also the highest point on an optimal mountain pass for the function $\underline{\sigma}_A$ between the corresponding eigenvalues. We use Algorithm 6.4 to find such points of coalescence, which are resolvent critical points.

To find the Wilkinson distance for a matrix A, we have to identify two components of the pseudospectra of A which first coalesce when ϵ increases. The pair of components that first coalesce may not be unique. Once we identify two eigenvalues in the respective components, the global mountain pass with the two eigenvalues being the two endpoints and for the function $\underline{\sigma}_A$ solves the Wilkinson problem. Note however that the identification of the eigenvalues and the global mountain pass problem are potentially difficult problems which we will not address in this chapter; in this work we consider only the local problem.

We should note that other approaches for the Wilkinson problem include [2], which uses a Newton type method for the same local problem, and [70].

6.7 Wilkinson's problem: Implementation and numerical results

We first use a convenient fast heuristic to estimate which pseudospectral components first coalesce as ϵ increases from zero, as follows. We construct the Voronoi diagram corresponding to the spectrum, and then minimize the function $\underline{\sigma}_A : \mathbb{C} \to \mathbb{R}$ over all the line segments in the diagram (a fast computation, as discussed in the comments on Step 1(b) below). We then concentrate on the pair of eigenvalues separated by the line segment containing the minimizer. This is illustrated in Example 6.27 below.

We describe implementation issues of Algorithm 6.4.

Step 1(a): Approximately minimizing the distance between a pair of points in distinct components seem challenging in practice, as we discussed briefly in Section

6.2. In the case of pseudospectral components, we have the advantage that computing the intersection between any circle and the pseudospectral boundary is an easy eigenvalue computation [71]. This observation can be used to to check optimality conditions or algorithm design for step 1(a). We note that in our numerical implementation, step 1(a) is never actually performed.

Step 1(b): Finding the global minimizer in step 1(b) of Algorithm 6.4 is easy in this case. Byers [26] proved that ϵ is a singular value of $A-(x+iy)I$ if and only if iy is an eigenvalue of

$$\begin{pmatrix} x-A^* & -\epsilon I \\ \epsilon I & A-x \end{pmatrix}.$$

Using Byer's observation, Boyd and Balakrishnan [17] devised a globally convergent and locally quadratic convergent method for the minimization problem over $\mathbb{R}$ of $y \mapsto \underline{\sigma}_A(x+iy)$. We can easily amend these observations to calculate the minimum of $\underline{\sigma}_A(x+iy)$ over a line segment efficiently by noticing that if $|z|=1$, then

$$\underline{\sigma}_A(x+iy) = \underline{\sigma}(A-(x+iy)I) = \underline{\sigma}(z(A-(x+iy)I)).$$

Example 6.27. We apply our mountain pass algorithm on the matrix

$$A = \begin{pmatrix} .461+.650i & .006+.625i & & & & \\ & .457+.983i & .297+.733i & & & \\ & & .451+.553i & .049+.376i & & \\ & & & .412+.400i & .693+.010i & \\ & & & & .902+.199i & \end{pmatrix}$$

The results of the numerical algorithm are presented in Table 6.1, and plots using EigTooL [91] are presented in Figure 6.6. We tried many random examples of bidiagonal matrices taking entries in the square $\{x+iy \mid 0 \le x,y \le 1\}$ of the same

form as A. The convergence to a critical point in this example is representative of the typical behavior we encountered.

In Figure 6.6, the top left picture shows that the first step in the Voronoi diagram method identifies the pseudospectral components corresponding to the eigenvalues $0.461 + 0.650i$ and $0.451 + 0.553i$ as the ones that possibly coalesce first. We zoom into these eigenvalues in the top right picture. In the bottom left diagram, successive steps in the bisection method gives better approximation of the saddle point. Finally in the bottom right picture, we see that the saddle point was calculated at an accuracy at which the level sets of $\underline{\sigma}_A$ are hard to compute.

There are other cases where the heuristic method fails to find the correct pair of eigenvalues whose components first coalesce.

Example 6.28. Consider the matrix A generated by the following Matlab code:

```
A=zeros(10);

A(1:9,2:10)= diag([0.5330 + 0.5330i, 0.9370 + 0.1190i,...
 0.7410 + 0.8340i, 0.7480 + 0.8870i, 0.6880 + 0.6700i,...
 0.2510 + 0.7430i, 0.9540 + 0.6590i, 0.2680 + 0.6610i,...
 0.2670 + 0.4340i]);

A=       A+diag([0.9850 + 0.7550i,0.8030 + 0.7810i,...
0.2590 + 0.5110i,0.3840 + 0.5310i,0.0080 + 0.5360i,...
0.9780 + 0.2720i,0.7190 + 0.3100i,0.5560 + 0.8370i,...
0.6350 + 0.7630i,0.5110 + 0.8870i]);
```

Table 6.1: Convergence data for Example 6.27. Significant digits are in bold.

i	$f(x_i)$	M_i	$\frac{M_i - f(x_i)}{f(x_i)}$	$\lvert x_i - y_i \rvert$
1	**6.1**325135002707E-4	**6.151109286**4335E-4	3.03E-03	5.23E-03
2	**6.151109**1521293E-4	**6.1511092861426**E-4	2.18E-08	1.40E-05
3	**6.1511092861422**E-4	**6.1511092861423**E-4	3.35E-15	9.97E-10

A sample run for this matrix is shown in Figure 6.7. The heuristic on minimal values of $\underline{\sigma}_A$ on the edges of the Voronoi diagram identifies the top left and central eigenvalues as a pair for which the pseudospectral components first coalesce. However, the correct pair should be the central and bottom right eigenvalues.

Here are a few more observations. In our trials, we attempt to find the Wilkinson distance for bidiagonal matrices of size 10×10 similar to the matrices in Examples 6.27 and 6.28. In all the examples we have tried, there was no need to perform step 1(a) of Algorithm 6.4 to achieve convergence to a critical point. The convergence for the matrix in Example 6.27 reflects the general performance of the (local) algorithm. As we have seen in Example 6.28, the heuristic for choosing a pair of eigenvalues may fail to choose the correct pseudospectral components which first coalesce as ϵ increases. In a sample of 225 runs, we need to check other pairs of eigenvalues 7 times. In such cases, a different choice of a pair of eigenvalues still gave convergence to the Wilkinson distance, though whether this must always be the case is uncertain. The upper bounds for the critical value are also better approximates of the critical values than the lower bounds.

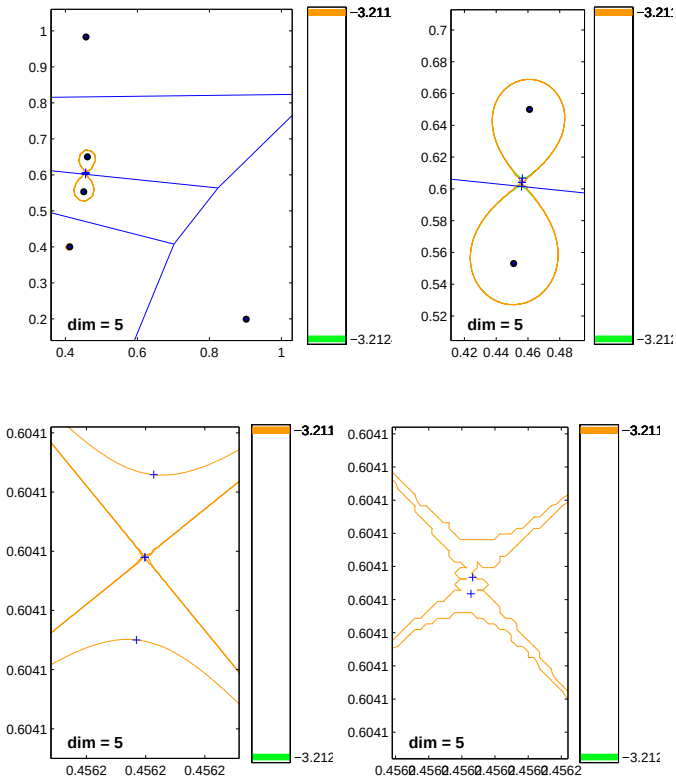


Figure 6.6: A sample run of Algorithm 6.4.

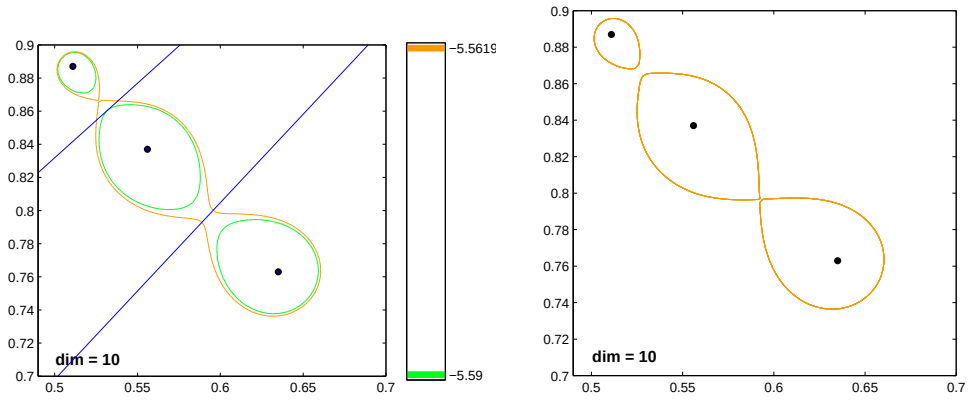


Figure 6.7: An example where the Voronoi diagram heuristic fails.

6.8 Non-Lipschitz convergence and optimality conditions

In this section, we discuss the convergence of Algorithm 6.1 in the non-Lipschitz case and give an optimality condition in step 2 of Algorithm 6.1. As one might expect in the smooth case in a Hilbert space, if x_i and y_i are closest points in the different components, $\nabla f(x_i) \neq \mathbf{0}$ and $\nabla f(y_i) \neq \mathbf{0}$, then we have

$$x_i - y_i = \lambda_1 \nabla f(y_i),$$
$$y_i - x_i = \lambda_2 \nabla f(x_i).$$

for $\lambda_1, \lambda_2 > 0$. The rest of this section extends this result to the nonsmooth case, making use of the language of variational analysis in the style of [80, 16, 31, 72] to describe the relation between subdifferentials of f and the normal cones of the level sets of f.

Closely related to the Fréchet normal cone is the proximal normal cone.

Definition 6.29. Let X be a Hilbert space and let $S \subset X$ be a closed set. If $x \notin S$ and $s \in S$ are such that s is a closest point to x in S, then any nonnegative multiple of $x - s$ is a *proximal normal vector* to S at s. The set of all proximal normal vectors is denoted $N_S^P(s)$.

The proximal normal cone and the Fréchet normal cone satisfy the following relation. See for example [16, Exercise 5.3.5].

Theorem 6.30. $N_S^P(\bar{x}) \subset \hat{N}_S(\bar{x})$.

Here is an easy consequence of the definitions.

Proposition 6.31. *Let S_1 be the component of $\operatorname{lev}_{\leq l_i} f$ containing x_0 and S_2 be the component of $\operatorname{lev}_{\leq l_i} f$ containing y_0. Suppose that x_i is a point in S_1 closest to*

S_2 and y_i is a point in S_2 closest to x_i. Then we have

$$(y_i - x_i) \in N^P_{\mathrm{lev}_{\leq l_i} f}(x_i) \subset \hat{N}_{\mathrm{lev}_{\leq l_i} f}(x_i).$$

Similarly, $(x_i - y_i) \in N^P_{\mathrm{lev}_{\leq l_i} f}(y_i)$. These are two normals of $\mathrm{lev}_{\leq l_i} f$ pointing in opposite directions.

The above result gives a necessary condition for the optimality of step 2 in Algorithm 6.1. We now see how the Fréchet normals relate to the subdifferential of f at x_i, y_i at $\bar{z}$.

It is clear from the definitions that the Fréchet subdifferential is contained in the limiting subdifferential, which is in turn contained in the Clarke subdifferential. Similarly, the Fréchet normal cone is contained in the limiting normal cone. We first state a theorem relating normal cones to subdifferentials in the finite dimensional case.

Theorem 6.32. *[80, Proposition 10.3] For a lsc function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, let $\bar{x}$ be a point with $f(\bar{x}) = \alpha$. Then*

$$\hat{N}_{\mathrm{lev}_{\leq \alpha} f}(\bar{x}) \supset \mathbb{R}_+ \hat{\partial} f(\bar{x}) \cup \{\mathbf{0}\}.$$

If $\partial f(\bar{x}) \not\ni \mathbf{0}$, then also

$$N_{\mathrm{lev}_{\leq \alpha} f}(\bar{x}) \subset \mathbb{R}_+ \partial f(\bar{x}) \cup \partial^\infty f(\bar{x}).$$

The corresponding result for the infinite dimensional case is presented below.

Theorem 6.33. *[16, Theorem 3.3.4] Let X be a Hilbert space and let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a lsc function. Suppose that $\liminf_{x \to \bar{x}} d(\hat{\partial} f(x); \mathbf{0}) > 0$ and $\xi \in \hat{N}_{\mathrm{lev}_{\leq f(\bar{x})} f}(\bar{x})$. Then, for any $\epsilon > 0$, there exist $\lambda > 0$, $(x, f(x)) \in \mathbb{B}_\epsilon((\bar{x}, f(\bar{x})))$ and $x^* \in \hat{\partial} f(x)$ such that*

$$|\lambda x^* - \xi| \leq \epsilon.$$

With these preliminaries, we now prove our theorem for the convergence of Algorithm 6.1 to a Clarke critical point.

Theorem 6.34. *Suppose that $f : X \to \mathbb{R}$, where X is a Hilbert space and f is lsc. If $\bar{z}$ is such that*

1. *$(\bar{z}, \bar{z})$ is a limit point of $\{(x_i, y_i)\}_{i=1}^{\infty}$ in Algorithm 6.1, and*
2. *f is continuous at $\bar{z}$.*

Then one of these must hold:

(a) $\bar{z}$ is a Clarke critical point,

(b) $\partial_C^{\infty} f(\bar{z})$ contains a line through the origin, or

(c) $\left\{\frac{y_i - x_i}{|y_i - x_i|}\right\}_i$ converges weakly to zero.

Proof. We present both the finite dimensional and infinite dimensional versions of the proof to our result.

Suppose the subsequence $\{(x_i, y_i)\}_{i \in J}$, where $J \subset \mathbb{N}$, is such that $\lim_{i \to \infty, i \in J}(x_i, y_i) = (\bar{z}, \bar{z})$. We can choose J so that none of the elements in $\{(x_i, y_i)\}_{i \in J}$ are such that $\liminf_{x \to x_i} d(\hat{\partial} f(x); \mathbf{0}) = 0$ or $\liminf_{y \to y_i} d(\hat{\partial} f(y); \mathbf{0}) = 0$, otherwise we have $\mathbf{0} \in \partial_C f(\bar{z})$ by the definition of the Clarke subdifferential, which is what we seek to prove. (In finite dimensions, the condition $\liminf_{x \to x_i} d(\hat{\partial} f(x); \mathbf{0}) = 0$ can be replaced by $\mathbf{0} \in \partial f(x_i)$.) We proceed to apply Theorem 6.33 (and Theorem 6.32 for finite dimensions) to find out more about $\hat{N}_{\mathrm{lev}_{\leq l_i} f}(x_i)$.

We first prove the result for finite dimensions. If $\mathbf{0} \in \partial f(\bar{z})$, we are done. Otherwise, by Proposition 6.31 and Theorem 6.32, there is a positive multiple of

$v = \lim_{i\to\infty} \frac{y_i - x_i}{|y_i - x_i|}$ that lies in either $\partial f(\bar{z})$ or $\partial^\infty f(\bar{z})$. Similarly, there is a positive multiple of $-v = \lim_{i\to\infty} \frac{x_i - y_i}{|y_i - x_i|}$ lying in either $\partial f(\bar{z})$ or $\partial^\infty f(\bar{z})$. If either v or $-v$ lies in $\partial f(\bar{z})$, then we can conclude $\mathbf{0} \in \partial_C f(\bar{z})$ from the definitions. Otherwise both v and $-v$ lie in $\partial_C^\infty f(\bar{z})$, so $\mathbb{R}\{v\} \subset \partial_C^\infty f(\bar{z})$ as needed.

We now prove the result for infinite dimensions. The point $\bar{z}$ is the common limit of $\{x_i\}_{i\in J}$ and $\{y_i\}_{i\in J}$. By the optimality of $|x_i - y_i|$ and Proposition 6.31, we have $y_i - x_i \in \hat{N}_{\mathrm{lev}_{\le l_i} f}(x_i)$ and $x_i - y_i \in \hat{N}_{\mathrm{lev}_{\le l_i} f}(y_i)$. By Theorem 6.33, for any $\kappa_i \to 0_+$, there is a $\lambda_i > 0$, $x_i' \in \mathbb{B}_{\kappa_i|x_i - y_i|}(x_i)$ and $x_i^* \in \hat{\partial} f(x_i')$ such that $|\lambda_i x_i^* - (y_i - x_i)| < \kappa_i |y_i - x_i|$. Similarly, there is a $\gamma_i > 0$, $y_i' \in \mathbb{B}_{\kappa_i|y_i - x_i|}(y_i)$ and $y \in \hat{\partial} f(y_i')$ such that $|\gamma_i y_i^* - (x_i - y_i)| < \kappa_i |x_i - y_i|$. If either x_i^* or y_i^* converges to $\mathbf{0}$, then $\mathbf{0} \in \partial_C f(\bar{z})$, and we are done. Otherwise, by the Banach Aloaglu theorem, the unit ball is compact, so $\left\{\frac{1}{|x_i^*|} x_i^*\right\}_i$ and $\left\{\frac{1}{|y_i - x_i|}(y_i - x_i)\right\}_i$ have weak cluster points. We now show that they must have the same cluster points by showing that their difference converges to $\mathbf{0}$ (in the strong topology). Now,

$$\begin{aligned}\left|\frac{\lambda_i x_i^*}{|y_i - x_i|}\right| &\le \left|\frac{\lambda_i x_i^*}{|y_i - x_i|} - \frac{y_i - x_i}{|y_i - x_i|}\right| + \left|\frac{y_i - x_i}{|y_i - x_i|}\right| \\ &\le \kappa_i + 1,\end{aligned}$$

and similarly, $1 - \kappa_i \le \left|\frac{\lambda_i x_i^*}{|y_i - x_i|}\right|$, so $\left|\frac{\lambda_i x_i^*}{|y_i - x_i|}\right| \to 1$, and thus

$$\left|\frac{\lambda_i x_i^*}{|y_i - x_i|} - \frac{x_i^*}{|x_i^*|}\right| = \left|\left|\frac{\lambda_i x_i^*}{|y_i - x_i|}\right| - \left|\frac{x_i^*}{|x_i^*|}\right|\right| \to 0.$$

This means that

$$\left|\frac{x_i^*}{|x_i^*|} - \frac{y_i - x_i}{|y_i - x_i|}\right| \le \left|\frac{\lambda_i x_i^*}{|y_i - x_i|} - \frac{x_i^*}{|x_i^*|}\right| + \left|\frac{\lambda_i x_i^*}{|y_i - x_i|} - \frac{y_i - x_i}{|y_i - x_i|}\right| \to 0,$$

which was what we claimed earlier. This implies that $\frac{x_i^*}{|x_i^*|}$ and $\frac{y_i^*}{|y_i^*|}$ have weak cluster points that are the negative of each other.

We now suppose that conclusion (c) does not hold. If $\{x_i^*\}_i$ has a nonzero weak cluster point, say $\bar{x}^*$, then $\bar{x}^*$ belongs to $\partial_C f(\bar{z})$. Then $\{y_i^*\}_i$ either has a

weak cluster point $\bar{y}^*$ that is strictly a negative multiple of $\bar{x}^*$, which implies that $\mathbf{0} \in \partial_C f(\bar{z})$ as claimed, or there is some $\bar{y}^{*,\infty} \in \partial_C^\infty f(\bar{z})$ which is a negative multiple of $\bar{x}^*$, which also implies that $\mathbf{0} \in \partial_C f(\bar{z})$ as needed.

If neither $\{x_i^*\}_i$ or $\{y_i^*\}_i$ converges weakly, then two (nonzero) weak cluster points of $\frac{x_i^*}{|x_i^*|}$ and $\frac{y_i^*}{|y_i^*|}$ that point in opposite directions give a line through the origin in $\partial_C^\infty f(\bar{z})$ as needed. □

In finite dimensions, conclusion (b) of Theorem 6.34 is precisely the lack of "epi-Lipschitzness" [80, Exercise 9.42(b)] of f. One example where Algorithm 6.1 does not converge to a Clarke critical point but to a point with its singular subdifferential $\partial_C^\infty f(\cdot)$ containing a line through the origin is $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = -\sqrt{|x|}$. Algorithm 6.1 converges to the point 0, where $\partial_C f(0) = \emptyset$ and $\partial_C^\infty f(0) = \mathbb{R}$. We do not know of an example where only condition (c) holds.

BIBLIOGRAPHY

[1] R. Alam & S. Bora. *On sensitivity of eigenvalues and eigendecompositions of matrices*, Linear Algebra and its applications 396 (2005) 273–301.

[2] R. Alam, S. Bora, R. Byers & M.L. Overton. *Characterization and construction of the nearest defective matrix via coalescence of pseudospectral components*, submitted, 2009.

[3] A. Ambrosetti. *Critical points and nonlinear variational problems*, Mémoires de la Société Mathématique de France, Sér. 2, 49 (1992), p. 1–139.

[4] A. Ambrosetti & P. Rabinowitz. *Dual variational methods in critical point theory and applications*, J. Funct. Anal., 14 (1973), pp. 349–381.

[5] H. Attouch & J. Bolte. *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Programming, **116** (2009), 5–16.

[6] J.-P. Aubin & I. Ekeland. *Applied Nonlinear Analysis*, Wiley 1984. Reprinted by Dover 2007.

[7] J.-P. Aubin & H. Frankowska. *Set-Valued Analysis*, Birkhauser, 1990.

[8] B. Bank, J. Guddat, D. Klatte, B. Kummer & K. Tammer. *Nonlinear Parametric Optimization*, Akademie-Verlag, Berlin, 1982.

[9] V. Barutello & S. Terracini. *A bisection algorithm for the numerical mountain pass*, Nonlinear differ. equ. appl. 14 (2007) 527–539.

[10] A. Ben-Tal & A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM, Philadelphia, 2001.

[11] R. Benedetti & J.-J. Risler. *Real algebraic and semi-algebraic sets*, Hermann, Paris, 1990.

[12] J. Bolte, A. Daniilidis, A.S. Lewis & M. Shiota. *Clarke critical values of subanalytic Lipschitz continuous functions.* Ann. Pol. Mat. 87 (2005) 13–25.

[13] J. Bolte, A. Daniilidis & A.S. Lewis. *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, *SIAM J. Optim.* **17** (2006), 1205–1223.

[14] J. Bolte, A. Daniilidis & A.S. Lewis. *Tame functions are semismooth*, Math. Programming *(*Series B*)* **117** (2009), 5–19.

[15] J.F. Bonnans & A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, NY, 2000.

[16] J.M. Borwein & Q.J. Zhu. *Techniques of Variational Analysis*, Springer, 2005.

[17] S. Boyd & V. Balakrishnan. *A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm*, Systems and Control Letters 15 (1990) 1–7.

[18] S. Boyd, L. El Ghaoui, E. Feron & V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM, Philadelphia, 1994.

[19] S. Boyd & L. Vandenberghe. *Convex Optimization*, Cambridge, 2004.

[20] A. Brown. *Functional dependence*, Trans. Amer. Math. Soc. **38** (1935), 379–394.

[21] J.V. Burke, A.S. Lewis, and M.L. Overton. *Optimization over pseudospectra*. SIAM Journal on Matrix Analysis 25:80–104, 2003.

[22] J.V. Burke, A.S. Lewis & M.L. Overton. *Optimization and pseudospectra, with applications to robust stability*. SIAM Journal on Matrix Analysis and its Applications, 25:80–104, 2003, Corrigendum: www.cs.nyu.edu/cs/faculty/overton/papers/pseudo_corrigendum.html

[23] J.V. Burke, A.S. Lewis & M.L. Overton. *Robust stability and a criss-cross algorithm for pseudospectra*. IMA Journal of Numerical Analysis (2003) **23**, 359–375.

[24] J.V. Burke, A.S. Lewis & M.L. Overton. *Convexity and Lipschitz behaviour of small pseudospectra*. SIAM Journal on Matrix Analysis and Applications, 29:586–595, 2007.

[25] J.V. Burke, A.S. Lewis & M.L. Overton. *Spectral conditioning and pseudospectral growth*. Numerische Mathematik, 107:27–37, 2007

[26] R. Byers. *A bisection method for measuring the distance of a stable matrix to the unstable matrices*, SIAM J. Sci. Stat. Comput., 9 (1988), pp. 875–881.

[27] F. Chaitin-Chatelin, A. Harrabi & A. Ilahi. *About Hölder condition numbers and the stratification diagram for defective eigenvalues*. Mathematics and Computers in Simulation, 54:397–402, 2000.

[28] Kung-Ching Chang. *Variational methods for non-differentiable functionals and their applications to partial differential equations*, Journal of Mathematical Analysis and its Applications, 80, 102–129 (1981).

[29] Y.S. Choi & P.J. McKenna. *A mountain pass method for the numerical solution of semilinear elliptic problems*, Nonlinear Anal., 20 (1993), pp. 417–437.

[30] G. Choquet. *Convergences*, Ann. Univ. Grenoble **23** (1948), 57–112.

[31] F.H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, New York, 1983. Republished as Vol. 5, Classics in Applied Mathematics, SIAM, 1990.

[32] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern & P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer-Verlag, New York, 1998.

[33] M. Coste. *An Introduction to O-minimal Geometry*, Instituti Editoriali e poligrafici internazionali, Universita di Pisa, 1999, available electronically at `http://perso.univ-rennes1.fr/michel.coste/`

[34] M. Coste. *An Introduction to Semialgebraic Geometry*, Instituti Editoriali e poligrafici internazionali, Universita di Pisa, 2002, available electronically at `http://perso.univ-rennes1.fr/michel.coste/`

[35] A. Daniilidis & C.H.J. Pang. *Continuity of set-valued maps revisited in the light of tame geometry*, preprint, 2009.

[36] M. Degiovanni & M. Marzocchi. *A critical point theory for nonsmooth functionals*, Ann. Math. Pura. Appl. 167 (1994), pp. 73–100.

[37] J.W. Demmel. *A counterexample for two conjectures about stability*, IEEE Trans. Auto. Control, AC-32:340–342, 1987.

[38] J.W. Demmel. *On condition numbers and the distance to the nearest ill-conditioned problem*, Numerische Mathematik, 51, 251–289, 1987.

[39] Z. Denkowska & J. Stasica. *Ensembles sous-analytiques à la polonaise*, Collection Travaux en Cours **69,** Hermann Mathmatiques, 2007.

[40] Zhonghai Ding, David Costa & Goong Chen. *A high-linking algorithm for sign-changing solutions of semilinear elliptic equations*, Nonlinear Analysis 38 (1999) 151–172.

[41] A.L. Dontchev & R.T. Rockafellar. *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Analysis, 12: 79–109, 2004.

[42] L. van den Dries. *Tame Topology and o-minimal Structures,* Cambridge, 1998.

[43] L. El Ghaoui & S.-I. Niculescu. *Robust decision problems in engineering: a linear matrix inequality approach*. In L. El Ghaoui and S.-I. Niculescu, editors, Advances in Linear Matrix Inequality Methods in Control, pages 3–37. SIAM, Philadelphia, 2000.

[44] M. Embree & L.N. Trefethen. *Pseudospectra Gateway* http://web.comlab.ox.ac.uk/projects/pseudospectra

[45] A. Göpfert, H. Riahi, C. Tammer & C. Zălinescu. *Variational Methods in Partially Ordered Spaces*, CMS Books in Mathematics **17,** Springer, 2003.

[46] G. Henkelman , G. Jóhannesson & H. Jónsson. *Methods for finding saddle points and minimum energy paths*, In: Progress in Theoretical Chemistry and Physics. S.D. Schwartz (ed.) Vol. 5, Kluwer 2000.

[47] J.-B. Hiriart-Urruty. *The deconvolution operation in convex analysis: an introduction*. Cybernetics Systems Analysis, 30:555–560, 1994.

[48] J. Horák. *Constrained mountain pass algorithm for the numerical solution of semilinear elliptic problems*, Numerische Mathematik 98 (2004) 251–276.

[49] R.A. Horn & C.R. Johnson. *Matrix Analysis*. Cambridge, 1985.

[50] R.A. Horn & C.R. Johnson. *Topics in Matrix Analysis*. Cambridge, 1991.

[51] P. Huard. *Background to point-to-set maps in mathematical programming*, Mathematical Programming study, 10(1979), pp 1–28.

[52] A. Ioffe. *Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems*, Proc. Amer. Math. Soc. **136** (2008), 3111–3119.

[53] A. Ioffe. *An invitation to tame optimization*, SIAM J. Optim. **19** (2008), 1894–1917.

[54] A.D. Ioffe & E. Scwhartzman. *Metric critical point theory 1: Morse regularity and homotopic stability of a minimum*, J. Math Pures Appl. 75 (1996), pp. 125–153.

[55] Youssef Jabri. *The Mountain Pass Theorem*, Cambridge, 2003.

[56] J. Jahn. *Vector Optimization. Theory, applications, and extensions.* Springer, 2004.

[57] M. Karow. *Geometry of Spectral Value Sets*. PhD Thesis, University of Bremen, 2003.

[58] M. Karow. *Eigenvalue condition numbers and a formula of Burke, Lewis and Overton*. Electronic Journal of Linear Algebra, 15:143–153, 2006.

[59] G. Katriel. *Mountain pass theorem and a global homeomorphism theorem*, Ann. Institut Henri Poincaré, Analyse Non Linéaire, 11 (1994), pp. 189–209.

[60] K. Kuratowski. *Les fonctions semi-continues dans l'espace des ensembles fermés*, Fundamenta Mathematicae **18** (1932), 148–159.

[61] A.S. Lewis. *Robust Regularization*, preprint, 2002.

[62] A.S. Lewis & C.H.J. Pang. *Variational analysis of pseudospectra*, SIAM J. Optim. **19** (2008), 1048–1072.

[63] A.S. Lewis & C.H.J. Pang. *Lipschitz behavior of the robust regularization*, Submitted to SIAM Journal on Control and Optimization (in second review), 2009.

[64] A.S. Lewis & C.H.J. Pang. *Level set methods for finding critical points of mountain pass type*, preprint, 2009

[65] W. Li. *Sharp Lipschitz constants for basic optimal solutions and basic feasible solutions of linear programs*, SIAM J. Control Optim., 32 (1994), pp. 140–153.

[66] Yongxin Li & Jianxin Zhou. *A minimax method for finding multiple critical points and its applications to semilinear PDES*, SIAM J. Sci. Comput., Vol 23, No. 3 , pp 840–865, 2001.

[67] Yongxin Li & Jianxin Zhou. *Convergence results of a local minimax method for finding multiple critical points*, SIAM J. Sci. Comput., Vol 24, No. 3, pp. 865–885, 2002.

[68] A.N. Malyshev. *A formula for the 2-norm distance from a matrix to the set of matrices with multiple eigenvalues*, Numer. Math. 83 (1999) 443-454.

[69] J. Mawhin & M. Willem. *Critical Point Theory and Hamiltonian Systems*, Springer, Berlin, 1989.

[70] E. Mengi. 2009. private communication.

[71] E. Mengi & M.L. Overton. *Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix*, IMA Journal of Numerical Analysis 25 (2005) pp. 648–669.

[72] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I and II*. Springer, Berlin, 2006.

[73] J.J. Moré & T. S. Munson. *Computing mountain passes and transition states*, Math. Program. Ser. B 100: 151–182 (2004).

[74] J. Munkres. *Topology* (2^{nd} edition) (Prentice Hall, 2000).

[75] L. Nirenberg. *Variational Methods in Nonlinear Problems. Topics in the calculus of variations (Montecatini Terme, 1987)*, 100-119, Lectures Notes in Mathematics, 1365, Springer, 1989.

[76] J. Oxtoby. *Measure and category. A survey of the analogies between topological and measure spaces* (2nd edition), Graduate Texts in Mathematics, (Springer, 1980).

[77] P.H. Rabinowitz. *Minimax Methods in Critical Point Theory with Applications to Differential Equations*, CBMS Regional Conference ser. Math, AMS, 65, 1986.

[78] S.M. Robinson. *Some continuity properties of polyhedral multifunctions*, Mathematical Programming Studies **14** (1981), 206–214.

[79] S.M. Robinson. *Solution continuity in monotone affine variational inequalities*, SIAM J. Optim. Volume 18, Issue 3, pp. 1046–1060 (2007)

[80] R.T. Rockafellar & R.J.-B. Wets. *Variational Analysis*. Springer, Berlin, 1998.

[81] A. Sard. *The measure of the critical values of differentiable maps*, Bull. Amer. Math. Soc. **48** (1942), 883–890.

[82] M. Struwe. *Variational Methods* (3^{rd} edition) (Springer, 2000).

[83] Shu-Chung Shi. *Semi-continuités génériques de multi-applications*, C.R. Acad. Sci. Paris (Srie A) **293** (1981), 27–29.

[84] S. Shi. *Ekeland's variational principle and the mountain pass lemma*, Acta. Math. Sin., (N.S.), 1, no. 4, 348–355 (1985).

[85] J.E. Sinclair & R. Fletcher. *A new method of saddle-point location for the calculation of defect migration energies*, J. Phys. C: Solid State Phys., pp 864–870, Vol 7, 1974.

[86] L.N. Trefethen & M. Embree. *Spectra and Pseudospectra*, Princeton, 2006.

[87] G.A. Watson. *Characterization of the subdifferential of some matrix norms*, Linear Algebra and its Applications, 170: 33-45, 1992.

[88] H. Whitney. *A function not constant on a connected set of critical points*, Duke Math. J. **1** (1935), 514–517.

[89] J.H. Wilkinson. *The Algebraic Eigenvalue Problem*, Oxford, 1965.

[90] M. Willem. *Un Lemme de déformation quantitatif en calcul des variations.* (French) [*A quantitative deformation lemma in the calculus of variations.*] Institut de Mathématiques pures et appliquées [Applied and Pure Mathematics Institute], Recherche de mathématiques [Mathematics Research] no. 19, Catholic University of Louvain, May 1992.

[91] T.G. Wright. *EigTool: a graphical tool for nonsymmetric eigenproblems*, 2002; available online at http://web.comlab.ox.ac.uk/pseudospectra/eigtool/

[92] Xudong Yao & Jianxin Zhou. *A local minimax characterization of computing multiple nonsmooth saddle critical points*, Math. Program., Ser. B 104, 749–760 (2005).

[93] Xudong Yao & Jianxin Zhou. *Unified convergence results on a minimax algorithm for finding multiple critical points in Banach spaces*, SIAM J. Num. Anal., 45 (2007) 1330–1347.